Generative Diagnosis Of Partial Discharge Fault Types In High-Voltage Cables Using Sparse Data Augmentation

Ren Peng*, Song Jun, Ji Hongwei, Ren Peng, Yang Yong, and Chen Jie

Shandong Luruan Digital Technology Co., LTD. No. 2008, Xinluo Street, High-tech Industrial Development Zone, Jinan City, Shandong Province, Yinhe Building, China

* *Corresponding author. E-mail: ren_peng001@outlook.com; renpieng@126.com*

Received: Dec. 24, 2025; Accepted: Apr. 01, 2026

Partial discharge (PD) in high-voltage cables often includes rare defect types, while fieldcollected pulse phase analysis (PRPD) spectra typically exhibit sparsity and class imbalance. Limited representative samples and weakly discriminative features hinder accurate extraction of deep time-frequency characteristics, leading to misclassification of defect types. To address this issue, this study proposes a generative diagnostic framework based on sparse data augmentation. An attention-enhanced BAGAN model is applied to augment sparse PRPD spectra, producing a balanced dataset across defect categories. The enhanced dataset is then processed using VMD-MSE to extract low-redundancy, high-discriminative time-frequency features. These features are subsequently input into an IKHA-optimized Deep Belief Network (IKHA-DBN) for defect classification. Experimental results show that generative augmentation increases rare defect samples by 700%, effectively eliminating class imbalance. The synthesized PRPD spectra exhibit strong consistency with genuine spectra in phase distribution and amplitude characteristics, confirming physical plausibility. In testing on 64 multi-class samples, only one misclassification occurred, demonstrating high diagnostic accuracy and sensitivity to minority defects. The results validate the robustness and effectiveness of the proposed method for intelligent diagnosis of rare PD defects in highvoltage cables.

Keywords: Sparse Data Augmentation, High-Voltage Cable, Partial Discharge, Defect Type Diagnosis, Generative Adversarial Network

http://dx.doi.org/10.6180/jase.202609_32.058

1. Introduction

High-voltage cables are extensively used in urban power grids and transmission systems, with their operational status being critical to power supply reliability and grid safety [1]. Partial discharge (PD), serving as an early indicator of cable insulation aging and defects, has become a key diagnostic method through pulse phase analysis (PRPD) for detection and analysis [2]. However, in practical applications, due to the scarcity of minority PD defect samples and complex detection conditions [3], PRPD data often exhibits sparsity and class imbalance. This hinders traditional diagnostic models from accurately identifying minority PD defects, thereby reducing diagnostic accuracy and reliability [4]. Therefore, how to effectively utilize foundational data with sparse and imbalanced characteristics for precise high-voltage cable partial discharge classification and defect type diagnosis has become a critical issue requiring urgent resolution in intelligent high-voltage cable diagnostics [5].

Currently, some scholars have conducted relevant research on cable partial discharge pattern recognition. Deep learning methods have been employed to extract PRPD spectral features from high-voltage cable PD signals, en-

abling defect classification and pattern recognition based on these features [6]. This approach depends on balanced datasets and may misclassify rare PD defects, while the Mycielski algorithm distinguishes defect types through temporal pattern matching and sequence similarity [7]. This approach is highly noise-sensitive and biased when rare PD defect samples are scarce, prompting the investigation of a DC-based diagnostic method for partial discharge defects in cross-linked polyethylene extruded cables [8]. Although partial discharge signal analysis enables defect classification and life assessment, it overlooks sparse low-probability defect data, limiting overall robustness [9].

The method ignores sparse imbalance, while BAGAN uses attention-based augmentation to generate realistic samples and improve performance [10]. VMD-MSE (Multiscale Entropy Variational Modal Decomposition) is employed for signal processing and time-frequency feature extraction. It decomposes complex signals into multiple modal components [11] It extracts multiscale discriminative features, uses DBNs for hierarchical learning, IKHA for optimized convergence, and a stacked autoencoder-SVM to improve detection accuracy and reduce false alarms [12].

Unlike prior GAN-based PD methods, the proposed BAGAN-VMD-MSE-IKHA-DBN framework enhances robustness and accuracy under sparse conditions. Based on the above analysis, this paper investigates a generative sparse data augmentation method for diagnosing partial discharge defect types in high-voltage cables.

2. Generative diagnostic method for partial discharge fault types in high-voltage cables

2.1. Generative Data Augmentation for Sparse High-Voltage Cable Partial Discharge Data

The sparse, imbalanced PRPD dataset biases models, while BAGAN mitigates this through class-balanced training and structured latent learning that preserves localized discharge patterns and avoids mode collapse [13], BAGAN leverages attention to preserve defect-specific PRPD patterns, generating realistic samples that fill sparse data gaps and enhance diagnostic reliability [14]. An attention-enhanced Conv4 model with weighted feature fusion improves class representation and similarity measurement, achieving better few-shot classification performance on miniImageNet and Omniglot [15].

A semi-supervised GAN model with revised loss functions improves faulty feeder identification under scarce labeled data by leveraging synthetic samples, with simulations and field tests validating its effectiveness [16]. An optimized GAN-based approach improves elevator fault diagnosis with limited labels by generating high-quality synthetic data, achieving accurate detection and reduced downtime [17]. This review compares traditional, signal-processing, and machine-learning methods for crane fault diagnosis and highlights future directions to improve safety and reliability [18].

This establishes a high-quality, balanced dataset for feature extraction and defect diagnosis, achieved by applying BAGAN to augment the original sparse partial discharge data. Attention in the BAGAN generator enhances spectral-phase features and highlights localized discharge regions, preserving sparse PRPD morphology while ensuring efficient, high-fidelity generation.

The overall architecture of the BAGAN model is shown in Fig. 1. In Fig. 1, Decon denotes the deconvolution layer,

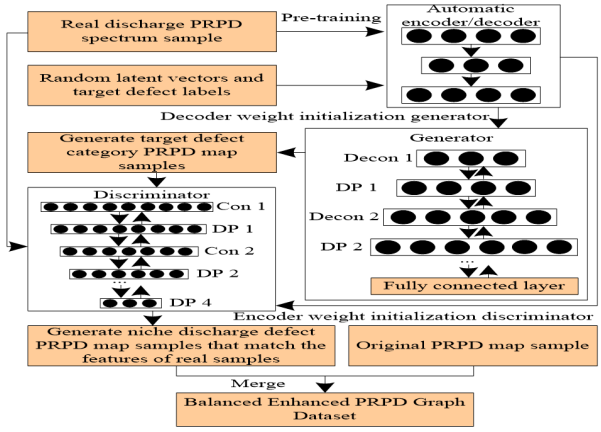


Fig. 1. Overall Architecture of BAGAN Model

DP represents the Dropout layer, and Con signifies the convolution layer. The training and sparse data generative augmentation process of the BAGAN model are as follows:

An attention-based encoder-decoder learns structured PRPD features and preserves defect morphology, improving rare defect generation and diagnostic robustness. Pre-trained generator and discriminator with gradient-penalized adversarial loss ensure stable PRPD generation, integrating authenticity, label matching, mislabel differentiation, and gradient penalty terms.

$$\begin{aligned}\chi^{(\varsigma)}(A_s, C, B_s, B_e, B_{\text{wrong}}) = &-\mathbb{E}_{(a_s,b_s)\sim(A_s,B_s)}\left[\log_2 \varsigma(a_s,b_s)\right]\\ &-\mathbb{E}_{(c,b_e)\sim(C,B_e)}\left[\log_2\left(1-\varsigma(\phi(c,b_e),b_e)\right)\right]\\ &-\mathbb{E}_{(a_s,b_{\text{wrong}})\sim(A_s,B_{\text{wrong}})}\left[\log_2\left(1-\varsigma(a_s,b_{\text{wrong}})\right)\right]\\ &+\chi_{\text{GP}}\end{aligned} \tag{1}$$

A_s and B_s denote real PRPD spectra and defect labels; a_s, b_s are real samples; b_e, b_{wrong} are generated and incorrect labels; $\phi(c, b_e)$ are generated samples, where $c \in C, \zeta$

and ϕ are the discriminator and generator, and χ_{GP} is the gradient penalty. χ_{GP} improves stability in distinguishing similar defects, enhancing model robustness.

The generator loss function aims to optimize the generation of partial discharge defect samples that are classified as genuine and match the target defect category. Its expression is:

$$L^{(\varphi)}(C, B_e) = -F\left[\log_2\left(\varsigma\left(\varphi\left(c, b_e\right)\right)\right)\right] \tag{2}$$

The trained generator produces realistic rare-defect PRPD spectra, and these are merged with original samples to create a balanced dataset for feature extraction and diagnosis.

2.2. Feature Extraction from the Enhanced Partial Discharge Phase-Frequency-Density

(PRPD) Dataset for High-Voltage Cables The balanced PRPD dataset contains phase, amplitude, and frequency information but is vulnerable to interference and equipment noise when directly used for defect diagnosis [19]. Therefore, time-frequency analysis is applied to extract more discriminative deep features [20]. VMD-MSE extracts noise-reduced, multiscale time-frequency features from balanced PRPD data, improving defect separability and diagnostic accuracy. Specifically, the VMD method decomposes the discharge signals corresponding to the PD PRPD spectra in this dataset into multiple physically meaningful intrinsic mode components, thereby separating useful discharge signals from noise [21, 22]. VMD-MSE features are classified by an IKHAoptimized DBN for accurate partial discharge defect diagnosis.

Let $X = \{x_1, x_2, \cdots, x_M\}$ denote the discharge signal corresponding to the PRPD spectrum dataset after generative augmentation, where M represents the length of this discharge signal. The decomposition process using the VMD method with this discharge signal as input is as follows:

(1) Formulate a Variational Constraint Problem Aiming to minimize the sum of modal bandwidths while ensuring the sum of modes equals the input discharge signal, the constrained problem is formulated as:

$$\begin{cases} \min_{\{w_l\},\{\rho_l\}} \left\{ \sum_{l=1}^{L} \left\| \partial_t \left[\left(\alpha(t) + \frac{\mathrm{j}}{\pi t} \right) w_l(t) \right] \mathrm{e}^{-\mathrm{j}\rho_l t} \right\|_2^2 \right\} \\ \text{s.t. } \sum_{l=1}^{L} w_l = x \end{cases} \tag{3}$$

Here, $\{w\}$ represents the L intrinsic modal components obtained by decomposing the input discharge signal via VMD; w_l denotes the l th intrinsic modal component, $l = 1, 2, \cdots, L$; ρ_l is the center frequency of the l th intrinsic modal component; x is the input discharge signal derived from the PRPD spectrum; and $\alpha(t)$ is the Dirac delta function, i.e., the unit impulse function.

(2) Converting the Variational Constraint Problem into an Unconstrained Problem

The constrained variational problem is transformed into an unconstrained one by introducing a quadratic penalty term β and a Lagrange operator $\delta(t)$, forming the extended Lagrangian.

$$\begin{aligned} l\left(\{w_l\}, \{\rho_l\}, \delta\right) = \beta \sum_{l=1}^{L} \left\| \partial_t \left[\left(\alpha(t) + \frac{\mathrm{i}}{\pi t} \right) w_l(t) \right] \mathrm{e}^{-\mathrm{i}\rho_l t} \right\|_2^2 \\ + \left\| x(t) - \sum_{l=1}^{L} w_l(t) \right\|_2^2 \\ + \left\langle \delta(t),\ x(t) - \sum_{l=1}^{L} w_l(t) \right\rangle \end{aligned} \tag{4}$$

(3) Iterative Solution ADMM iteratively finds the saddle point of the extended Lagrangian, obtaining the optimal solution and intrinsic mode components wlw_lwl with their center frequencies.

$$\begin{cases} \widehat{w}_l^{m+1}(\rho) = \dfrac{\hat{x}(\rho) - \sum_{i \neq l}^{L} \widehat{w}_i(\rho) + \frac{\hat{\delta}(\rho)}{2}}{1 + 2\beta\left(\rho - \rho_l\right)^2} \\ \rho_l^{m+1} = \dfrac{\int_0^{\infty} \rho \left| \widehat{w}_l^m(\rho) \right|^2 d\rho}{\int_0^{\infty} \left| \widehat{w}_l^m(\rho) \right|^2 d\rho} \end{cases} \tag{2}$$

Here, m is the iteration index; $\hat{w}_l^m(\rho)$ and $\hat{w}_l^{m+1}(\rho)$ denote the frequency-domain forms of the l-th mode w_l : $\hat{w}_i(\rho)$ is the i-th mode; $\hat{x}(\rho)$ is the input signal; and $\hat{\delta}(\rho)$ is the frequency-domain Lagrangian operator. Its update equation $\hat{\delta}(\rho)$ is computed iteratively.

$$\hat{\delta}^{\dot{m}+1}(\rho) \leftarrow \hat{\delta}^{\dot{m}}(\rho) + \tau \left[\hat{x}(\rho) - \sum_{l=1}^{L} \widehat{w}_l^{\dot{m}+1}(\rho) \right] \tag{6}$$

Here, τ is the ADMM step size controlling the update of $\hat{\delta}$,iteration stops when $\sum_{l=1}^{L} \| \hat{w}_l^{m+1} - \hat{w}_l^m \|_2^2 / \| \hat{w}_l^m \|_2^2 < \sigma$.

The output yields L time-domain modes $w_l(t)$, with reduced mode mixing, and multiscale entropy quantifies their complexity to highlight defect-specific differences.

(1) Parameter initialization: Set embedding dimension $= 2$, similarity tolerance $\kappa = 0.1\varepsilon$, and scale factor $\lambda = [1, 2, \ldots, \lambda_{\max}]$. Where, ε represents the standard deviation of the original discharge signal, and the scale factor λ must be adjusted to accommodate the signal's fluctuation characteristics.

(2) Coarsening: Coarsen the decomposed modal components $w_l(t)$ of the discharge signal according to the scale λ to obtain a new sequence:

$$y_j(\lambda) = \frac{1}{\lambda} \sum_{l=(j-1)\lambda+1}^{j\lambda} w_l \quad \left(1 \le j \le M'/\lambda\right) \tag{7}$$

Where, M' denotes the length of the modal componentw $w_l(t)$.

(3) Calculate multiscale entropy: For each new sequence obtained after coarse-graining, construct an -dimensional vector:

$$Y(l) = [y_l(\lambda), y_{l+1}(\lambda), \ldots, y_{l+n-1}(\lambda)] \tag{8}$$

Where, $1 \le l \le M' - n$. Based on the set similarity tolerance $\kappa = 0.1\varepsilon$, count the number of distances between constructed vectors that satisfy $D(l, j) < \kappa$, and calculate the proportion of this count relative to the total number of distances $M' - n + 1$. This proportion is defined as $\eta_l^n(\kappa)$, and its mean is defined as $\bar{\eta}^n(\kappa)$. Repeating this process yields $\bar{\eta}^{n+1}(\kappa)$. Based on this, the sample entropy of the new sequence $y_j(\lambda)$ is calculated as:

$$\text{SampEn}\left(n, \kappa, M'\right) = -\ln\left[\bar{\eta}^{n+1}(\kappa)/\bar{\eta}^n(\kappa)\right] \tag{9}$$

Integrating the sample entropies of sequences at different scales yields the MSE of the decomposed modal component $w_l(t)$ of the discharge signal:

$$\text{MSE}\left(w_l\right) = \text{SampEn}(y(\lambda), n, \kappa) \tag{10}$$

MSE evaluates structural complexity across scales, independent of amplitude; normalization and adaptive tolerance ensure stability, while VMD's mode isolation enables entropy features to capture defect-specific morphology for structure-based classification.

Combining the MSE of all modal components corresponding to each partial discharge PRPD spectrum into a feature vector $Z = \{\text{MSE}_{w_1}, \text{MSE}_{w_2}, \ldots, \text{MSE}_L\}$, followed by PCA dimensionality reduction (to avoid the curse of dimensionality), ultimately yields a lowredundancy, high-discriminative time-frequency feature vector z' for high-voltage cable partial discharge signals. This serves as input for the subsequent IKHA-DBN diagnostic model.

2.3. IKHA-DBN-Based Diagnosis of Partial Discharge Fault Types in High-Voltage Cables

Low-redundancy VMD-MSE features are classified using an IKHA-optimized DBN to achieve accurate partial discharge defect diagnosis.

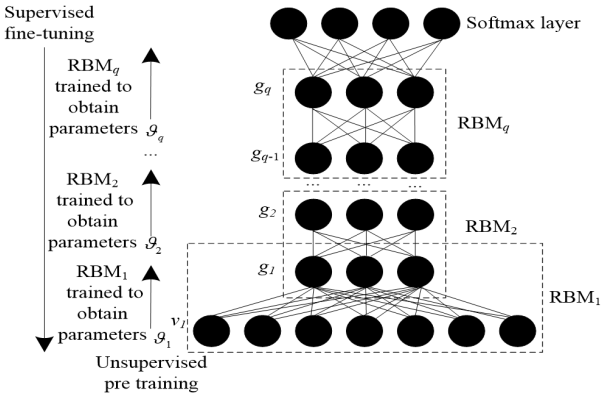


Fig. 2. Location of the KVMRT project

2.3.1. *Basic DBN Model Architecture and Training Logic*

The DBN stacks RBMs with pre-training and fine-tuning, using VMD-MSE features as input and a Softmax layer for partial discharge defect classification, is illustrated in Fig. 2.

The training steps for the DBN model are:

(1) Unsupervised pre-training: The core of each RBM layer in the DBN model is the energy function, defined as the energy relationship between the visible layer and hidden layer states:

$$P_\vartheta(v, g) = -\sum_{i=1}^{K_v} \mu_{vi} v_i - \sum_{j=1}^{K_g} \mu_{gj} g_j - \sum_{i=1}^{K_v} \sum_{j=1}^{K_g} v_i \omega_{ij} g_j \tag{11}$$

where v_i represents the visible layer neuron states (input features); g_j denotes the hidden layer neuron states; K_v and K_g are the numbers of visible and hidden layer neurons, respectively; μ_{vi} and μ_{gj} are the visible and hidden layer biases, respectively; ω_{ij} is the connection weight between the two layers; and $\vartheta = \left[\omega = (\omega_{ij})_{K_v \times K_g},\ \mu_v = (\mu_{vi})_{K_v},\ \mu_g = (\mu_{gj})_{K_g}\right]$ denotes the set of parameters for each RBM layer.

Through this energy function, the joint probability distribution of the RBM can be defined as:

$$R_\vartheta(v, g) = \frac{\exp\left(-P_\vartheta(v, g)\right)}{\zeta_\vartheta} \tag{12}$$

Where, ζ_ϑ is the normalization factor, which serves to normalize the probabilities. Its expression is:

$$\zeta_\vartheta = \sum_v \sum_g \exp\left(-P_\vartheta(v, g)\right) \tag{13}$$

In an RBM, hidden units independently activate via a sigmoid function based on visible inputs.

$$R_\vartheta\left(g_j=1\mid v\right)=\text{sigmoid}\left(\mu_{gj}+\sum_{i=1}^{K_v}\omega_{ij}v_i\right) \tag{14}$$

Where, $\text{sigmoid}\left(z'\right)=\left(1+\exp\left(z'\right)\right)^{-1}$ is the activation function, which maps input feature sample values to the range $(0,1)$ to adapt to the probability value interval. Similarly, given the known state of hidden layer neurons, the probability of activating visible layer neurons is obtained as:

$$R_\vartheta\left(v_i=1\mid g\right)=\text{sigmoid}\left(\mu_{vi}+\sum_{j=1}^{K_g}\omega_{ji}g_j\right) \tag{15}$$

(2) The contrast divergence (CD) algorithm is employed to optimize the RBM parameters ϑ, yielding their optimal values. The RBM parameters of the DBN model are randomly initialized. The time–frequency features of the partial discharge signal, extracted using VMD–MSE, serve as the initial visible layer neurons v^0. Based on Eq. (14), the initial hidden layer neurons g^0 are computed. Subsequently, the visible layer neurons v^1 are reconstructed according to Eq. (15). Then, using v^1, the hidden layer neurons g^1 are recalculated. The update equations for each model parameter are given as follows:

$$\begin{cases}\Delta\omega_{ij}=v\left(\left\langle v_i^0g_j^0\right\rangle-\left\langle v_i^1g_j^1\right\rangle\right)\\ \Delta\mu_{vi}=v\left(\left\langle v_i^0\right\rangle-\left\langle v_i^1\right\rangle\right)\\ \Delta\mu_{gj}=v\left(\left\langle g_j^0\right\rangle-\left\langle g_j^1\right\rangle\right)\end{cases} \tag{16}$$

Where, $\langle\cdot\rangle$ denotes the mathematical expectation; v represents the learning rate of the CD algorithm.

Given the large volume of partial discharge signal feature samples input, random mini-batch sampling is required during actual RBM training to enhance efficiency. Assuming the RBM samples M'' partial discharge signal feature samples per training iteration, the update equation for the RBM parameter ϑ at iteration q is:

$$\begin{cases}\omega_{ij,q}=\omega_{ij,q-1}+v\left[\sum_{m''=1}^{M''}\left(v_{i,m''}^{q-1}g_{j,m''}^{q-1}-v_{i,m''}^{q}g_{j,m''}^{q}\right)/M''\right]\\ \mu_{vi,q}=\mu_{vi,q-1}+v\left[\sum_{m''=1}^{M''}\left(v_{i,m''}^{q-1}-v_{i,m''}^{q}\right)/M''\right]\\ \mu_{gj,q}=\mu_{gj,q-1}+v\left[\sum_{m''=1}^{M''}\left(g_{j,m''}^{q-1}-g_{j,m''}^{q}\right)/M''\right]\end{cases} \tag{17}$$

(3) Supervised fine-tuning: A Softmax layer is added atop the pre-trained network. The backpropagation algorithm is employed to fine-tune the overall network parameters [23, 24], optimizing the classification accuracy of partial discharge defects.

2.3.2. IKHA-Based DBN Model Optimization

After RBM pre-training, IKHA optimizes DBN parameters via diagnostic error minimization, improving convergence stability and defect discrimination.

The position update formula for the leader krill in the IKHA algorithm is:

$$\xi_j^1=\begin{cases}\xi_j^{\text{best}}+o_1\cdot\left(o_2\cdot\left(\bar{\xi}_j-\underline{\xi}_j\right)+\underline{\xi}_j\right),o'\geq0.5\\ \xi_j^{\text{best}}-o_1\cdot\left(o_2\cdot\left(\bar{\xi}_j-\underline{\xi}_j\right)+\underline{\xi}_j\right),o'<0.5\end{cases} \tag{18}$$

Where ξ_j^{best} denotes the current optimal parameter combination; $\bar{\xi}_j$ and $\underline{\xi}_j$ represent the upper and lower bounds of the parameters, respectively; o_2 and o' are random numbers within the interval $[0,1]$; and $o_1=2e^{-\left(4\frac{l}{L}\right)}$ balances global search and local exploration, where l is the current iteration count and L is the maximum number of iterations.

(3) Optimization process:

IKHA encodes DBN parameters as krill individuals and optimizes them via diagnostic error minimization to obtain the final IKHA-DBN model.

2.3.3. Defect Type Diagnosis Implementation Based on the IKHA-DBN Model

The IKHA-optimized DBN classifies VMD-MSE-extracted time-frequency features to perform high-voltage cable partial discharge defect diagnosis, as illustrated in Fig. 3.

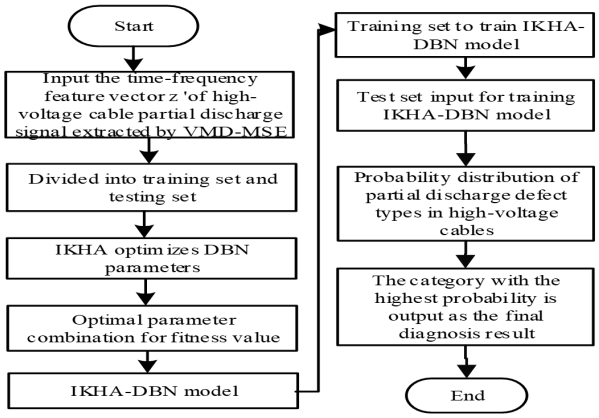


Fig. 3. Process Diagram of Partial Discharge Defect Type Diagnosis for High-voltage

3. Cables based on ikha-dbn model

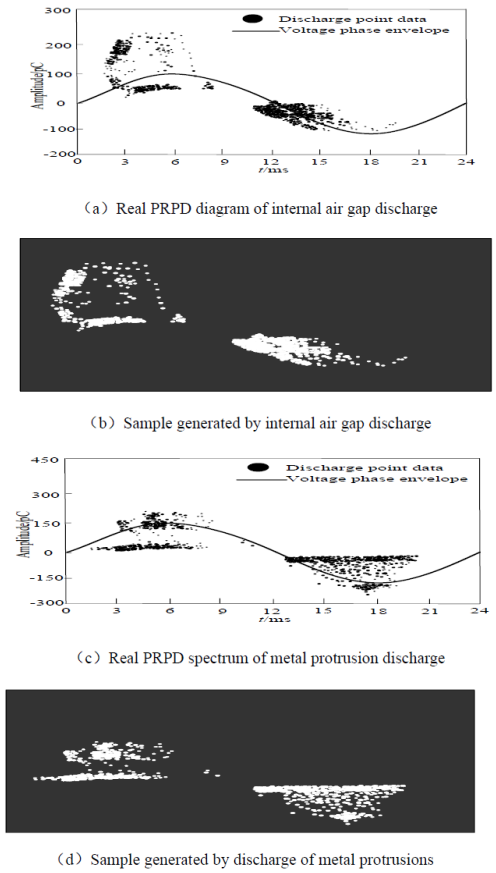
4. Analysis of experimental results

IKHA-DBN classified 70/30-split features, validated on a 50 kV XLPE cable using Wuhan Hezhong PRPD data, Its key parameters are shown in Table 1.

Table 1. Key parameters of experimental partial discharge detector

Parameter category	Parameter values / descriptions
Rated capacity	50 kVA
Input power	380 V, 50 Hz, single-phase
Output voltage range	0–150 kV (RMS, continuously adjustable)
Rated output current	Approximately 333 mA (50 kVA/150 kV)
Voltage accuracy	0.50%
Output	50 Hz (power frequency sine wave)
System partial discharge level	$\leq$ 3 pC
Short-circuit impedance	$\leq$ 5%
No-load current	$\leq$ 10%
System noise level	$\leq$ 65 dB
Control method	Fully automatic computer control, supporting PRPD spectrum acquisition and analysis
Typical applicable scenarios	Cables and transformers up to 50 kV, GIS systems; power-frequency withstand voltage and partial discharge testing in accordance with GB/T 7354 and DL/T 1576

A 50 kV cable setup produced PRPD data for five defect types, and generative augmentation was used to correct class imbalance (Table 2). Table 2 shows augmentation created a balanced 2,000-sample dataset, increasing rare defects by 700% with realistic synthetic PRPD data.



（a）Real PRPD diagram of internal air gap discharge

（b）Sample generated by internal air gap discharge

（c）Real PRPD spectrum of metal protrusion discharge

（d）Sample generated by discharge of metal protrusions

Fig. 4. Real Defect PRPD spectrum and Generated Defect Samples

Figure 4 shows generated samples closely match real rare defects, and the optimized IKHADBN achieves near-perfect accuracy after 700% minority augmentation.

5. Methods

Figure 5 shows IKHA-DBN achieved near-perfect classification with only one misdiagnosis, outperforming other methods due to balanced augmentation, VMD-MSE features, and IKHA optimization.

6. Conclusion

This study proposes a sparse data augmentation-based generative diagnostic framework for high-voltage cable partial discharge classification. An attention-enhanced BAGAN balances rare defect samples, VMD-MSE extracts discriminative time-frequency features, and IKHA optimizes the DBN to avoid local optima and improve convergence. The integrated approach achieves high diagnostic accuracy with only one misclassification, demonstrating strong robustness and effectiveness for rare defect detection under imbalanced data conditions.

7. Declarations

Funding: This work was supported by the Science and Technology Program of Hebei Province (Grant No. E2025502049), and by the Science and Technology Project of Shandong Luruan Digital Technology Co., LTD., entitled "Development of Domestic High-Voltage Cable Partial Discharge and Grounding Circulating Current Intrinsic Temperature Measurement Device" (Project No. B90630250009).

Conflicts of interests: Authors do not have any conflicts.

Data Availability Statement: The PRPD dataset, including original and augmented spectra with labels and features, is available from the corresponding author upon

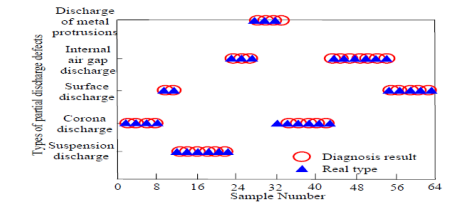
Table 2. Sample size of PRPD spectra data for various types of partial discharge defects before and after generative enhancement

Defect Type	Original sample quantity/piece	Number of enhanced samples/piece
Corona discharge	150	400
Surface discharge	150	400
Tip discharge	150	400
Internal air gap discharge	50	400
metal protrusion discharge	50	400
Total/piece	550	2000

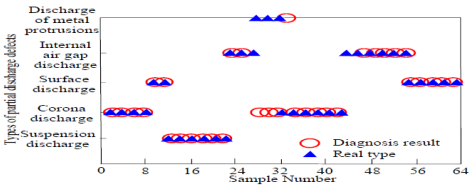
reasonable academic request with stated intended use and citation agreement.

Code availability: Not applicable.

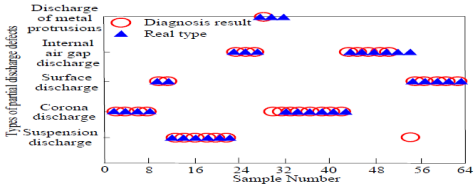
Authors' Contributions: Ren Peng, Song Jun, Ji Hongwei, is responsible for designing the framework, analyzing the performance, validating the results, and writing the article. Yang Yong, Chen Jie is responsible for collecting the information required for the framework, provision of software, critical review, and administering the process.



（a）The diagnostic results of the method described in this article



（b）Diagnosis results of deep learning algorithms



（c） Mykielski's algorithm diagnosis results

Fig. 5. Comparison of Partial Discharge Defect Type Diagnosis Results of Various

References

[1] O. Arikan, C. C. Uydur, and C. F. Kumru, (2023) *"Insulation Evaluation of MV Underground Cable with Partial Discharge and Dielectric Dissipation Factor Measurements"* **Electric Power Systems Research *220***: 109338. DOI: 10.1016/j.epsr.2023.109338.

[2] M. Rajamayil and V. Basharan, (2025) *"A Novel Semi-Supervised Power Transformer Defect Monitoring Technique Using Unreliable Pseudo-Labels with Highly Imbalanced Partial Discharge Signals"* **Electrical Engineering *107***(4): 4939–4957. DOI: 10.1007/s00202-024-02793-y.

[3] P. Ranjan, M. Oancea, Q. Han, F. O. Bahdad, C. Onoufriou, and M. Seltzer-Grant, (2024) *"Partial Discharge Monitoring of C3F7CN/CO2 Mixture Retrofilled in Gas Insulated Busbar"* **IEEE Transactions on Power Delivery *39***(6): 3212–3222. DOI: 10.1109/TPWRD.2024.3454440.

[4] I. A. Devi, R. V. Maheswari, and R. Rajesh, (2023) *"Recognition of Fused Partial Discharge Patterns in High Voltage Insulation Systems: A Hybrid DCNN and SVM Based Approach"* **IETE Journal of Research *69***(10): 7553–7568. DOI: 10.1080/03772063.2022.2038702.

[5] S. Mishra, P. P. Singh, and P. C. Bordin, (2024) *"Diagnostics Analysis of Partial Discharge Events of the Power Cables at Various Voltage Levels Using Ramping Behavior Analysis Method"* **Electric Power Systems Research *227***: 109988. DOI: 10.1016/j.epsr.2023.109988.

[6] M. H. Saad, S. Hashima, A. I. Omar, M. M. Fouda, and A. Said, (2025) *"Deep Learning Approach for Cable Partial Discharge Pattern Identification"* **Electrical Engineering *107***(2): 1525–1540. DOI: 10.1007/s00202-024-02571-w.

[7] F. Serttas and F. O. Hocaoglu, (2025) *"A Novel Partial Discharge Signal Detection and Estimation Method: Mycielski Algorithm"* **Electrical Engineering *107***(3): 2829–2843. DOI: 10.1007/s00202-024-02656-6.

[8] M. Fikri and Z. Abdul-Malek, (2023) *"Partial Discharge Diagnosis and Remaining Useful Lifetime in XLPE Extruded Power Cables under DC Voltage: A Review"* **Electrical Engineering *105***(6): 4195–4212. DOI: 10.1007/s00202-023-01935-y.

[9] J. M. Park, J. Y. Kang, Y. C. Moon, J. Lee, and J. C. Jeon, (2023) *"Analysis on the Partial Discharge Characteristics of 22.9kV CN-CV Power Cable Termination Defect"* **Transactions of the Korean Institute of Electrical Engineers *72***(11): 1546–1553. DOI: 10.5370/KIEE.2023.72.11.1546.

[10] N. V. Edward, A. Jenefa, T. M. Thiyagu, A. Lincy, and A. Taurshia, (2024) *"DeepGAN: Utilizing Generative Adversarial Networks for Improved Deep Learning"* **International Journal of Knowledge and Intelligent Engineering Systems *28***(4): 732–748. DOI: 10.3233/KES-230326.

[11] R. K. Mandal, H. Devarajan, S. Nandi, T. B. Shanker, and M. Ghassemi, (2025) *"An Underdetermined Single-Channel Blind-Source-Separation for Multisource Acoustic-Emission-Based Partial Discharge Signals in Power Transformers"* **IEEE Transactions on Plasma Science *53***(4): 788–797. DOI: 10.1109/TPS.2025.3538746.

[12] N. S. Allur, (2020) *"Phishing Website Detection Based on Multidimensional Features Driven by Deep Learning: Integrating Stacked Autoencoder and SVM"* **Journal of Science Technology *5***(6):

[13] M. Carioni, A. I. José, and D. Walter, (2025) *"Extremal Points and Sparse Optimization for Generalized Kantorovich–Rubinstein Norms"* **Foundations of Computational Mathematics *25***(1): 103–144. DOI: 10.1007/s10208-023-09634-7.

[14] S. K. Hamed, M. J. A. Aziz, and M. R. Yaakub, (2025) *"A Data Augmentation Approach Based on Various GAN Models to Address Class Imbalance in Fine-Grained Multimodal Fake News Datasets"* **Computing *107***(1): 1–43. DOI: 10.1007/s00607-025-01413-2.

[15] K. G. Kim and B. T. Lee, (2025) *"Graph Structure Based Data Augmentation Method"* **Biomedical Engineering Letters *15***(2): 283–289. DOI: 10.1007/s13534-024-00446-4.

[16] X. Meng, X. Wang, S. Yin, and H. Li, (2023) *"Few-shot Image Classification Algorithm Based on Attention Mechanism and Weight Fusion"* **Journal of Engineering and Applied Science *70***(1): 14. DOI: 10.1186/s44147-023-00186-9.

[17] X. Yu, J. Gu, X. Zhang, and J. Mao, (2023) *"GAN-based Semi-Supervised Learning Method for Identification of the Faulty Feeder in Resonant Grounding Distribution Networks"* **International Journal of Electrical Power & Energy Systems *144***: 108535. DOI: 10.1016/j.ijepes.2022.108535.

[18] X. Lv, Z. Lu, Z. Huang, and Z. Wei, (2025) *"Improving Fault Diagnosis in Elevator Systems with GAN-based Synthetic Data"* **International Journal of Sensor Networks *47***(1): 26–35. DOI: 10.1504/IJSNET.2025.143899.

[19] Y. Li, J. Liu, X. Tang, J. Pan, W. Liu, Y. Huang, and Z. Li, (2025) *"Fault Diagnosis Methods for Electromechanical Special Equipment: Review and Prospects"* **Measurement Science and Technology *36***(7): 076115. DOI: 10.1088/1361-6501/adeacf.

[20] D. Kim, S. Park, and S. B. H. Hwang, (2023) *"Fair Classification by Loss Balancing via Fairness-Aware Batch Sampling"* **Neurocomputing *518***: 231–241. DOI: 10.1016/j.neucom.2022.11.018.

[21] R. Li, Z. Zhou, S. Liu, and T. Wang, (2024) *"Simulation of Online Monitoring Method for Partial Discharge at High-Voltage Power Cable Terminals"* **Computer Simulation *41***(4): 80–84.

[22] J. Yeo, H. Jin, A. R. Mor, C. Yuen, N. Pattanadech, and W. Tushar, (2023) *"Localisation of Partial Discharge in Power Cables through Multi-Output Convolutional Recurrent Neural Network and Feature Extraction"* **IEEE Transactions on Power Delivery *38***(1): 177–188. DOI: 10.1109/TPWRD.2022.3183588.

[23] V. Souza, G. Xavier, J. Cruz, A. Oliveira, A. Serres, and P. Lazaridis, (2023) *"The Practical Application of Bio-Inspired PMA for the Detection of Partial Discharges in High Voltage Equipment"* **Sensors *23***(23): 9307. DOI: 10.3390/s23239307.

[24] W. Hassan, M. Shafiq, M. Choudhary, I. Palu, and G. A. Hussain, (2023) *"Investigating the Progression of Insulation Degradation in Power Cable Based on Partial Discharge Measurements"* **Electric Power Systems Research *221***: 1–12. DOI: 10.1016/j.epsr.2023.109452.