IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 16 July 2024, accepted 12 August 2024, date of publication 19 August 2024, date of current version 28 August 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3445974

RESEARCH ARTICLE

Semi-Supervised Learning-Based Partial Discharge Diagnosis in Gas-Insulated Switchgear

HO TRONG TAI[1], YOUNG-WOO YOUN[2,3], (Associate Member, IEEE), HYEON-SOO CHOI[4], AND YONG-HWA KIM[1], (Member, IEEE)

[1]Department of Computer Science and Information, Korea National University of Transportation, Uiwang, Gyeonggi 16106, Republic of Korea
[2]Smart Grid Research Division, Korea Electrotechnology Research Institute, Gwangju 61751, Republic of Korea
[3]Kim Jaecul Graduate School of AI, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea
[4]Genad Systems, Naju, Jeollanam 58296, Republic of Korea

Corresponding author: Yong-Hwa Kim (yongkim@ut.ac.kr)

This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry & Energy (MOTIE) of the Republic of Korea under Grant 20225500000120 and Grant 20221A10100011.

ABSTRACT Effective monitoring and diagnosis of partial discharge (PD) in power equipment are crucial for maintenance, particularly given the expectations of significant increases in energy generation and consumption. Although deep neural networks have been widely applied in PD fault detection and classification, their performance is hindered by insufficient labeled data available for power equipment. This study proposes a semi-supervised learning (SSL) method to address the scarcity of labeled training data for PD classification in gas-insulated switchgear (GIS). The proposed SSL was validated based on phase-resolved PD and on-site noise using an ultra-high frequency (UHF) PD measurement system. Experimental results show that the proposed SSL achieves a high classification accuracy of 94.59% by effectively utilizing unlabeled data to enhance classification performance in GIS.

INDEX TERMS Semi-supervised learning (SSL), fault diagnosis, phase-resolved partial discharge (PRPD), gas-insulated switchgear (GIS).

I. INTRODUCTION

Gas-Insulated switchgear (GIS) has become increasingly prevalent in power systems, playing a critical role in ensuring the reliability of the overall grid infrastructure. Its significance lies in its maintenance-free nature, minimal environmental impact, compact design, and high reliability [1]. However, GIS is often subjected to harsh operating conditions characterized by high temperatures and pressures, making it prone to insulation defects. Additionally, inherent risks during manufacturing, transportation, and assembly stages can lead to latent faults [2].

The common insulation defects include metal tip defects, particle defects, floating electrode defects, corona defects, and void defects. They are recognized as some of the most destructive and undesirable phenomena that, in the long run, leads to progressive local deterioration of the insulation system and may result in a complete insulation breakdown [3], [4], [5], [6]. The most prevalent fault type among these is the floating-related fault, which arises from poor contact, loose bolts, or rusted components in GIS [7]. The most common explanation for the appearance of PDs is the aging process of the insulation system, which is usually caused by cracks, voids, impurities, and other imperfections within the insulation materials [8]. In an electrical power system, most critical failures within the electrical power distribution system are partially or fully associated with PDs [9]. To locate PD sources and recognize fault types, the UHF method is one of the most promising, given its relative ease of application, particularly under on-site conditions. Due to its sensitivity and robust anti-interference capabilities, the UHF method has been widely used and has proves to be an effective and cost-effective solution [10]. In this study, the UHF method is employed in a PD measurement system to monitor the real-time conditions of GIS and diagnose existing defects.

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaodong Liang.

Several studies have applied machine learning techniques to classify PDs based on UHF methods, achieving excellent performance in PD pattern-recognition classification. The time-resolved partial discharge (TRPD) mode [11], [12], [13] involves analyzing the shape of the discharge pulse using time-domain features and frequency-domain features, and a combination of both to examine thoroughly the PD characteristics. The phase-resolved partial discharge (PRPD) mode [14], [15], [16] involves an analysis of the phase-amplitude-number pattern of the PRPD. This mode identifies defect shapes by analyzing the number of PD pulses and the maximum or average amplitude in each phase of the PRPD. Compared with TRPD, PRPD has the advantages of mature technology, good stability, and a small amount of data, making it widely used. However, these methods exhibit certain shortcomings, such as relying heavily on expert experience, introducing the potential for artificial errors due to manual interventions, features extracted by diverse algorithms lacking shareability and transferability, and an extended duration of training [17].

Recent advancements in deep neural networks (DNNs) have revolutionized the processing of large datasets, overcoming the limitations of traditional machine learning methods in PD classification in GIS. Recurrent neural networks (RNNs) with long short-term memory (LSTM) have been employed to classify PDs in GIS using one-power-cycle PRPD sequences [18]. Convolutional neural networks (CNNs) have demonstrated automatic feature extraction capabilities for power transformer PD classification using UHF signals [19]. A previous study explored the combination of CNN and LSTM models has been explored to extract valuable information from UHF PRPDs for PD classification [20]. A stacked sparse auto-encoder (SSAE) was used to assess PD severity using raw UHF PRPD data as input in GIS [21]. However, numerous parameters in DNNs pose a risk of overfitting [22]. In addition, many of these algorithms rely on supervised learning, necessitating fully labeled datasets, which can be time-consuming and labor-intensive [23]. Furthermore, supervised learning approaches may overlook potentially valuable data features by not discovering and clustering them independently.

To mitigate the scarcity of training data, novel approaches such as self-supervised learning (Self-SL) [24], [25], [26], [27] and semi-supervised learning (SSL) [28], [29], [30], [31] have been proposed to harness unlabeled data and learn essential features with minimal supervision. Unlike Self-SL methods, which operate in an entirely unsupervised manner, SSL functions in scenarios where both a substantial volume of unlabeled data and a limited amount of labeled data are available. SSL enables neural networks to extract richer data features with fewer labels, samples, or iterations, thereby significantly reducing the need for manual labeling and data preparation.

In this study, we propose an SSL for fault diagnosis using PRPDs in GIS. The proposed SSL not only deals with limited labels but also diminishes the overfitting problem. The model trained on labeled data with cross-entropy loss is used to predict the labels for unlabeled samples after data augmentation is performed. Predicted labels with the highest confidence are used as pseudo-labels. The model is then trained with unlabeled data in a supervised using the generated pseudo-labels. Through a combination of consistency regularization and curriculum learning, the proposed model efficiently utilizes high-accuracy pseudo-labels, simultaneously mitigating the effects of low-accuracy pseudo-labels, thereby reducing the overall loss.

The contributions of this study are as follows.

- The SSL with the pseudo-labeling is applied for the first time to classify faults in GIS. In the proposed SSL, we use two data augmentation techniques for PRPDs to address the challenge of limited labeled data in GIS fault diagnosis.
- The performance of the proposed SSL is verified using small labeled data through on-site noise and PRPD experiments, where PRPD data include four types of faults such as corona, floating, particle and void. The proposed SSL outperforms Self-SL by 1.8% and achieves a classification accuracy of 94.59% for on-site noise and PD data in GIS.

The remainder of this study is organized as follows. We briefly introduce the experimental data for GIS in Section II. Section III describes the problem description and the proposed SSL. Performance evaluations are presented in Section IV. The study is concluded in Section V.

II. FAULT SIMULATION EXPERIMENT

This section presents an overview of the experimental setup and an analysis of the PRPD and on-line noise using a UHF sensor to evaluate the PD characteristics in GIS.

A. EXPERIMENTAL SYSTEM

Fig. 1 illustrates the experimental system designed for PRPD analysis. Fig. 1a presents a block diagram of the measurement system, including a high-voltage source, voltage divider, artificial PD cell, external UHF sensor, and data acquisition system (DAS), while Fig. 1b shows the high-voltage test site. The high-voltage source operated at a frequency of 60 Hz, supplying high voltage to the test subjects, while the voltage divider reducing the voltage across them. The voltage was increased incrementally until the PD signal became evident, and the discharge intensity was adjusted within a specified range. The voltage gradually decreased to zero by the end of the test. As shown in Fig. 1c, the DAS comprised of a band-pass filter, logarithmic amplifier, peak detector, analog-to-digital converter (ADC), and PC. The PD signals from the UHF sensor were filtered through a band-pass filter to limit their frequency. Low-amplitude pulses from the sensor output passed through a logarithmic amplifier, facilitating the detection of PD pulses.

In this study, the PD measurements with UHF sensors utilized a 45 dB logarithmic amplifier with a bandwidth

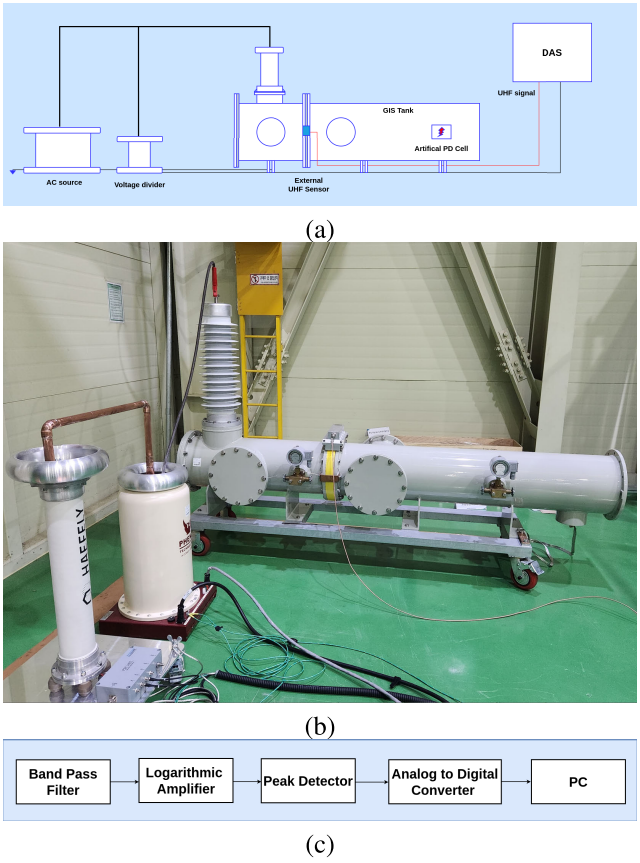

FIGURE 1. An experimental system for PRPD analysis: (a) Block diagram, (b) High-voltage test site, and (c) Data Acquisition System (DAS).

between 500M Hz and 1.5 GHz. The peak detector identifies the peak value of the UHF PD pulses. After the peak detector, the DAS used the ADC with 1024 x f_m samples per second, where f_m = 60Hz is the power frequency. The maximum value was then captured at every 8 samples in the DAS and P = 128 samples per power cycle were used for the PRPD measurements.

As presented in [18], artificial cells were used to simulate for four types of PD: corona, floating, void and particle. The artificial cells were filled with 0.2 *MPa* SF6 gas. To simulate the corona discharge, a sharp protrusion was affixed to an electrode, engineered to create a hyper-localized electric field enhancement through a needle, with a very minute 10-μm tip radius and 1-*mm* diameter. The needle was positioned at a distance of 10 *mm* away from the grounded electrode, with a test voltage of 11 *kV*. To replicate the scenario of an isolated or unconnected cell, the fabricated floating electrode cell was subjected to a test voltage of 10 *kV*. The distance between the middle electrode and the high-voltage electrode was set to 10 *mm*, while the distance between the middle electrode and the grounded electrode was maintained at 1 *mm*. For the artificial void defect, a small gap was present between the epoxy disc and upper electrode, mirroring and magnifying the intricacies of the void discharge at a test voltage of 8 *kV*. The particle discharge was recreated by precisely placing a 1-*mm*-diameter sphere above a concave ground electrode. A high voltage of 10 *kV* was used in the testing process, connected to a larger 45-*mm*-diameter sphere fixed at a distance of 10 *mm* from the grounded electrode.

B. PRPD AND ONLINE NOISE ANALYSIS

Using the DAS, the measured signal for M = 3600 power cycles is defined in matrix form as

$$\mathbf{X} = \begin{bmatrix} x(1,1) & x(1,2) & \dots & x(1,P) \\ . & . & \dots & . \\ . & . & \dots & . \\ . & . & \dots & . \\ x(M,1) & x(M,2) & \dots & x(M,P) \end{bmatrix} \quad (1)$$

where $x(m,p) \in \{0,1,\dots,255\}$ denotes the range of the measured signal at the p^{th} data point of the m^{th} power cycle.

Fig 2 shows the time-domain representations of PRPD patterns of the four types of faults in GIS, obtained using UHF sensors over 3600 power cycles. Each data point on the 3D graph corresponds to a PD pulse, representing its phase angle and amplitude. Additionally the number of PDs per 3600 power cycles is illustrated by employing different colors for overlapping points in phase amplitude 2D density plots. For the corona, as shown in Fig. 2a, PDs are distinctly observed in the negative half-cycle. This is because the corona PDs are predominantly observed in the negative half-cycle with around 270° at inception and spread out above inception [32], [33]. In Fig. 2b, for the floating fault, PDs are clearly observed in both the positive and negative halves and are centralized with high amplitudes at 60° and 200°. As shown in Fig. 2c, the majority of PDs for void faults are present in the positive half-cycle of the applied voltage. In Fig. 2d, PDs for the particle fault are concentrated within a low-level range of the measured signal.

Noise was measured using an online UHF PD monitoring system for on-site GIS at substations in South Korea, where the noise was measured using the same system with the other faults as in Fig. 2. For noise measurements, PD monitoring systems are installed on-site and noise is measured for GISs under normal conditions. The DAS uses an 8-bit ADC to store raw data, and the amplitude of noise is quantified as an integer from 0 to 255. As presented in Fig. 3, it can be seen that the pulses of noise signals are scattered without regularity in all ranges of phases and power cycles.

Fig. 4 shows statistical features using *mean*(PDs) and *max*(PDs) for on-site noise and PRPD measurements, where $mean(\text{PDs}) = mean(\mathbf{X}) = \frac{\sum_{m=1}^{M}\sum_{p=1}^{P} x(m,p)}{MP}$ and $max(\text{PDs}) = max(\mathbf{X}) = max\{x(1,1), x(1,2), \dots, x(M,P)\}$. For on-site noise, the *mean*(PDs) ranges from 0.1 to 1.2, while the *max*(PDs) varies between 28 and 90. In the case of corona, the *mean*(PDs) ranges from 0.1 to 2.2, with the *max*(PDs) falling within the ranges of 15 to 51, though numerous samples exhibited outliers extends up to 161. The *max*(PDs) for voids and particles showed a similar range, from around 17 to over 240. The *mean*(PDs) for voids ranges from 0.1 to 1.2, while

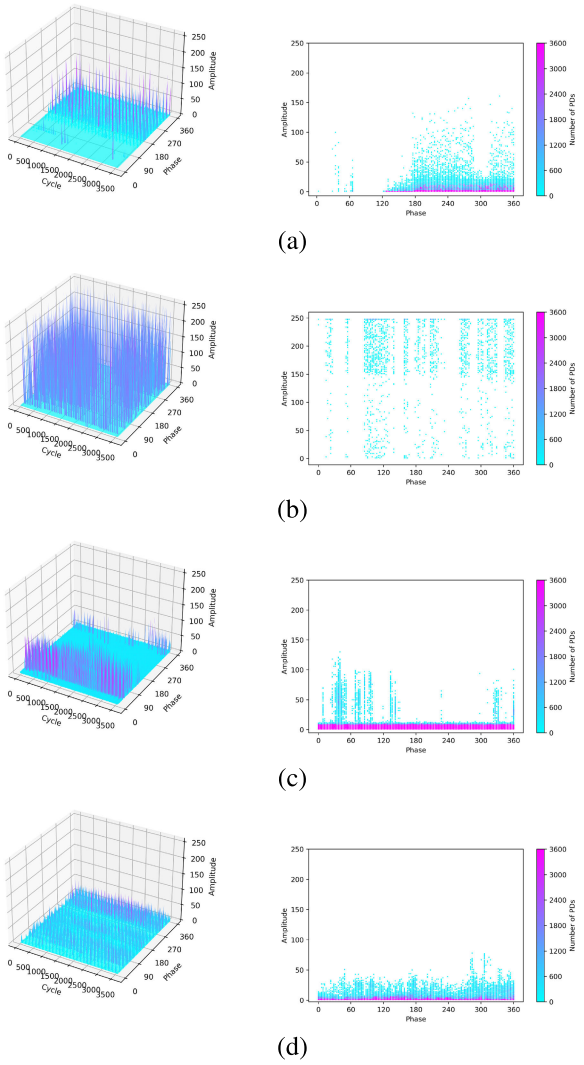

FIGURE 2. Sequential phase-resolved PDs corresponding phase-amplitude density plots for four types of faults in GIS: (a) Corona, (b) Floating, (c) Void and (d) Particle. In phase-amplitude density plots, the "Number of PDs" axis represents the total count of PDs occurring across phase and amplitude over 3600 power cycles.

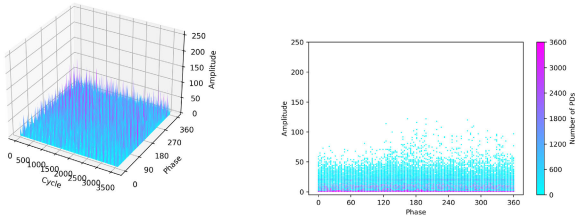

FIGURE 3. Sequential phase-resolved on-site noise corresponding phase-amplitude density plots. In phase-amplitude density plots, the "Number of PDs" axis represents the total count of pulses of noise signals occurring across phase and amplitude over 3600 power cycles.

for particles, it extended up to 2.2. Most floating samples have a *max*(PDs) range of 250, with *mean*(PDs) spanning from 0.3 to 2.2, and some samples have outliers reaching up to 5.6. Therefore, distinguishing between noise and PRPDs is

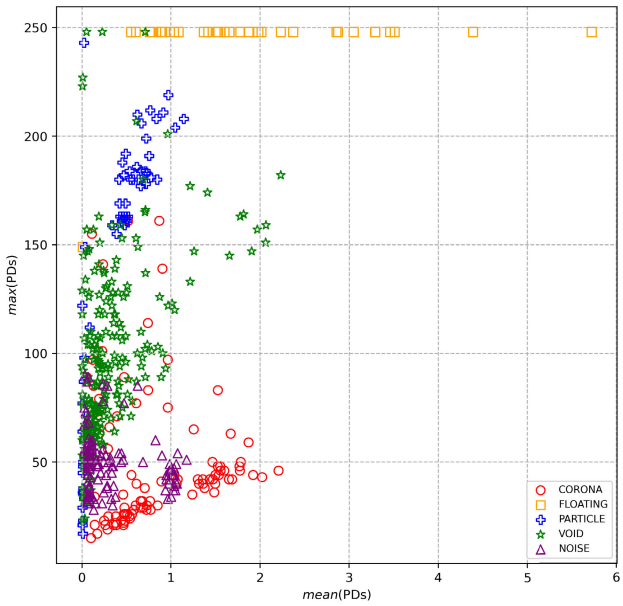

FIGURE 4. Statistical features for on-site noise and PRPD measurements using *mean*(PDs) and *max*(PDs).

challenging due to information loss in the feature extraction process by statistical parameters such as *mean*(PDs) and *max*(PDs).

Table 1 shows a statistical summary of on-site noise and PRPD measurements, utilizing both *max*(PDs) and *mean*(PDs). The floating data type has the highest minimum value for the minimum of *max*(PDs), while the other types are similar to noise levels. For the maximum of *max*(PDs), all fault types exceed noise levels. When considering the mean of *mean*(PDs), floating shows the highest value, as shown in Fig. 2b and Fig. 4, whereas void and particle values are comparable to noise. Additionally, the standard deviation of *mean*(PDs) for floating is the highest, with void, particle, and noise having standard deviations around 0.3.

III. PROPOSED SCHEME

A. PROBLEM DESCRIPTION

The deep learning-based fault diagnosis approaches require plenty of labeled datasets to learn the feature representations in embedding feature space [22]. Therefore, the lack of labeled training data significantly affects the accuracy of deep learning-based fault classification in GIS. Also, obtaining the labeled training data through fault simulation experiments can be time-consuming and labor-intensive. In this study, a model trained on labeled data was used to predict the labels for unlabeled data. The predicted pseudo-labels with higher confidence were subsequently used as labels for unlabeled data. The model was then trained with unlabeled data in a supervised setting using the generated pseudo-labels. Our goal was to improve the confidence level of the classifier to improve the classification accuracy.

TABLE 1. Statistical summary of on-site noise and PRPD measurements in GIS.

Type	Corona	Floating	Void	Particle	Noise
Min of max(PDs)	15	149	23	17	28
Max of max(PDs)	161	248	248	243	90
Mean of $mean$(PDs)	0.833	1.920	0.323	0.362	0.344
Standard deviation of $mean$(PDs)	0.558	1.189	0.398	0.333	0.363

For the notation, we use X, Y, Z to denote minibatches of samples, labels, and logits, respectively. Our data set can be divided into two parts: $X_L \subset \mathbb{R}^{n\times d}$ and $X_U \subset \mathbb{R}^{m\times d}$ denote minibatches of labeled samples and unlabeled samples, respectively, where where n is the minibatch size for labeled samples, m is the size of minibatches for unlabeled samples, and d is the input dimension (for PRPDs $d = M \cdot P$). The minimatch of labels $Y_L^{(i)} \subset \{0, 1\}^{n\times h}$ is corresponding to Y_L. In addition, we use $Y^{(i)}$ to refer the i-th row, and $Y^{(i,j)}$ to refer to the (i, j)-th element of Y. Model $f(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}^h$ takes samples as the input and outputs logits for each class of h classes. Our problem was a classification task using PRPDs in GIS.

B. PROPOSED SEMI-SUPERVISED LEARNING

Fig. 5 shows a block diagram of the proposed SSL. The proposed SSL consists of data augmentation, distribution alignment, relative confidence threshold, pseudo-labeling, and curriculum learning. Here, the distribution alignment helps constrain the distribution of the class predictions to align more with the true distribution [34]. A relative confidence threshold is utilized to address the issue of poor calibration in machine-learning models [35]. Consistency regularization utilizes unlabeled data based on the assumption that the model outputs similar predictions when fed perturbed versions of the same sample [36]. Curriculum learning can provide performance improvements over the standard training approach based on random data shuffling, without any additional computational costs [37].

For PRPDs, we use two data augmentation techniques: scale and crop [38]. Scale augmentation $\alpha(\cdot)$ is a function that multiplies each element of the input by a random variable n with a Gaussian distribution $N(1, 0.01)$. The purpose of the scale strategy is to examine the variations in the magnitude of partial discharges in relation to the severity of the faults. Crop augmentation $\beta(\cdot)$ is a function that sets a group of elements with a randomly probability of $c\%$ of the input matrix to be zero.

In a supervised fashion, scale augmentation is applied to each labeled sample in the batch before being processed through the model to obtain the logits as

$$Z_L = f(\alpha(X_L); \theta), \tag{2}$$

where θ denotes the set of parameters of the proposed SSL. The loss for a batch is defined by l_s and is calculated by taking the average of cross-entropy loss for each sample as

$$l_s = \frac{1}{n}\sum_{i=1}^{n} H(Y_L^{(i)}, Z_L^{(i)}), \tag{3}$$

where $H(p(x), q(x)) = -\sum p(x) \log q(x)$ denotes the regular cross-entropy used for classification.

Pseudo-labeling involves applying scale augmentation to unlabeled data and selecting entries with predicted probabilities above a specified confidence threshold. These selected entries serve as pseudo-labels, which are then compared with the output of the model on the crop-augmented data. Each unlabeled sample is subjected two augmentations and fed into the model to compute the logits as

$$Z_U = f(\alpha(X_U); \theta) \tag{4}$$

and

$$Z'_U = f(\beta(X_U); \theta). \tag{5}$$

The confidence threshold is then obtained by

$$\gamma = \frac{\tau}{n}\sum_{i=1}^{n} \underset{j\in[1\ldots h]}{max}(\hat{Y}_L^{(i,j)}), \tag{6}$$

where τ is the parameter for the confidence threshold and $\hat{Y}_L$ is the model's predicted class distribution as

$$\hat{Y}_L = softmax(Z_L) \in \mathbb{R}^{n.h}. \tag{7}$$

The predicted labels for the unlabeled data can then be obtained after applying the distribution alignment as

$$\tilde{Y}_U = normalize(\hat{Y}_U \frac{\mathbb{E}[\hat{Y}_L]}{\mathbb{E}[\hat{Y}_U]}) \in \mathbb{R}^{m.h} \tag{8}$$

where $normalize(\cdot)$ ensures that the distribution still sums to 1, and

$$\hat{Y}_U = softmax(Z_U) \in \mathbb{R}^{m.h} \tag{9}$$

and $\mathbb{E}[\hat{Y}_L], \mathbb{E}[\hat{Y}_U] \in \mathbb{R}^h$ are the expected of model's predicted distribution. The pseudo-labels are represented by the binary mask $\delta \in \{0, 1\}^m$

$$\delta^{(i)} = \underset{j\in[1\ldots h]}{max}(\tilde{Y}_U^{(i,j)}) \geq \gamma, \tag{10}$$

Consistency regularization involves crop augmented unlabeled data, and its output is then compared with our pseudo-label to compute the cross-entropy loss. The total unlabeled batch loss is represented by l_u and is given by

$$l_u = \frac{1}{m}\sum_{i=1}^{m} H(sg(\tilde{Y}_U^{(i)}), Z'^{(i)}_U).\delta^{(i)} \tag{11}$$

where entries for which the confidence is less than γ are not contributed to l_u loss. Because guess labels from unlabeled data can lead to unstable training and can hinder convergence, $sg(\cdot)$ is the stop gradient function used to

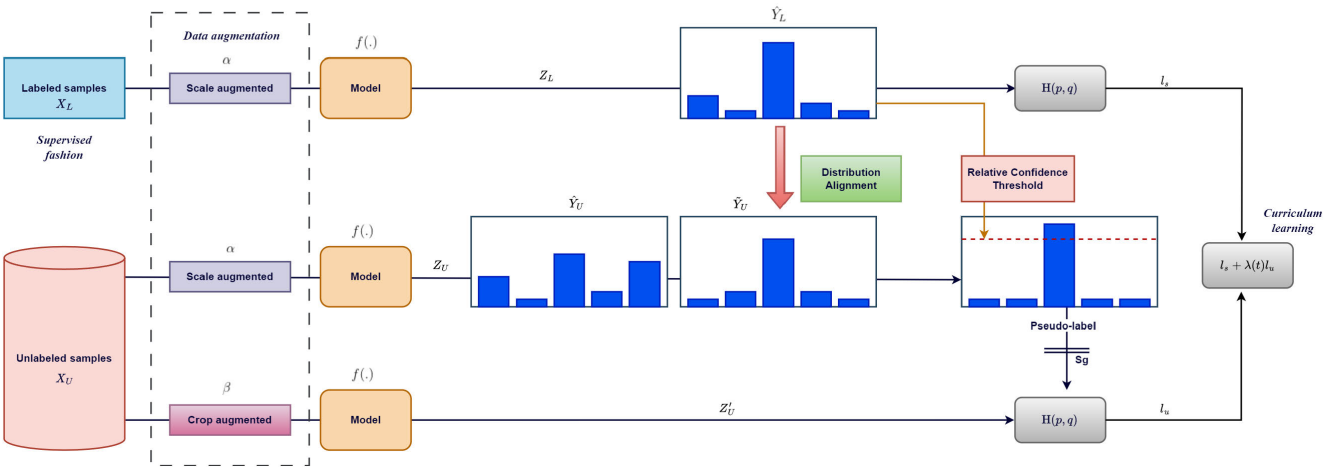

FIGURE 5. Our proposed semi-supervised learning.

TABLE 2. The number of samples for training and test sets.

Type	Training	Test	Total
Corona	33	14	47
Floating	13	5	18
Void	104	45	149
Particle	23	10	33
Noise	83	37	120
Unlabeled	368	-	368
Total	624	111	735

mitigate back-propagation from guess labels to improve convergence and the over all performance of the model training [30]. This approach leverages additional information from unlabeled data to enhance the model's performance.

Curriculum learning involves the final combination of two losses to obtain a total loss that is optimized to improve the model and is defined as

$$l = l_s + \lambda(t) l_u, \tag{12}$$

where t is the index for the epoch and $\lambda(t)$ is a function to balance the importance of labeled and unlabeled data during the training process and facilitate faster convergence. The Adam optimizer is used to optimize the learnable parameters during the training process [40]. An overview of the training process is presented in Algorithm 1.

IV. PERFORMANCE EVALUATIONS

A performance evaluation of the proposed SSL using PRPD and on-site noise measurements was next conducted. Table 2 shows the number of samples for the training and test sets, where the training set consisted of 256 labeled and 368 unlabeled samples, and the test set had 111 labeled samples. For labeled samples, on-site noise and four types of PRPDs, namely, corona, floating, void, and particle, were considered. All simulations were performed on a personal computer with NVIDIA GeForce RTX 2080 Ti GPU and 128 Gb RAM.

Algorithm 1 Our Proposed SSL Algorithm

Input: Labeled batch X_L, unlabeled batch X_U, one-hot label batch Y_L, scale augmentation $\alpha(\cdot)$, crop augmentation $\beta(\cdot)$, parameter for confidence threshold τ, unlabeled weight ratio $\lambda(t)$, model $f(\cdot)$ taking samples as input and outputting logits for each of h classes, E epochs.

for $t = 1 to E$ **do**
  for $i = 1 to n$ **do**
    $Z_L^{(i)} = f(\alpha(X_L^{(i)}); \theta)$
    for $j = 1 to h$ **do**
      $\hat{Y}_L^{(i,j)} = \frac{e^{Z_L^{(i,j)}}}{\sum_{k=1}^{h} e^{Z_L^{(i,k)}}}$
    end for
  end for
  $\gamma = \frac{\tau}{n} \sum_{i=1}^{n} \max_{j \in [1...h]} (\hat{Y}_L^{(i,j)})$
  $\mathbb{E}[\hat{Y}_L] = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{h} \hat{Y}_L^{(i,j)}$
  $l_s = \frac{1}{n} \sum_{i=1}^{n} H(Y_L^{(i)}, Z_L^{(i)})$
  for $i = 1 to m$ **do**
    $Z_U^{(i)} = f(\alpha(X_U^{(i)}); \theta)$
    $Z_U'^{(i)} = f(\beta(X_U^{(i)}); \theta)$
    for $j = 1 to h$ **do**
      $\hat{Y}_U^{(i,j)} = \frac{e^{Z_U^{(i,j)}}}{\sum_{k=1}^{h} e^{Z_U^{(i,k)}}}$
    end for
  end for
  $\mathbb{E}[\hat{Y}_U] = \frac{1}{m} \sum_{i=1}^{m} \sum_{j=1}^{h} \hat{Y}_U^{(i,j)}$
  $\tilde{Y}_U^{(i,j)} = \text{normalize}(\hat{Y}_U^{(i,j)} \cdot \frac{\mathbb{E}[\hat{Y}_L]}{\mathbb{E}[\hat{Y}_U]})$
  $\delta^{(i)} = \max_{j \in [1...h]} (\tilde{Y}_U^{(i,j)}) \geq \gamma$
  $l_u = \frac{1}{m} \sum_{i=1}^{m} H(sg(\tilde{Y}_U^{(i)}), Z'^{(i)}_U) \cdot \delta^{(i)}$
  return $l = l_s + \lambda(t) l_u$
  update θ using ADAM
end for
return θ

Output: Our proposed SSL method

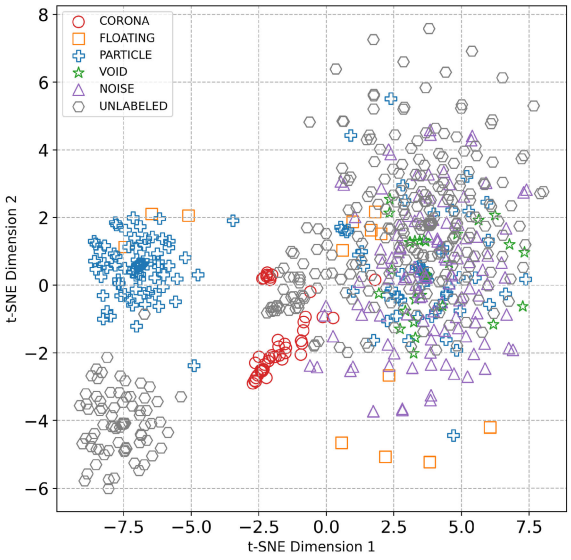

FIGURE 6. t-SNE visualization.

Fig. 6 shows the t-distributed stochastic neighbor embedding representation (t-SNE) of all samples to provide intuitive insight into the structural composition of the data set [41]. With a reduction in the dimensions of the data, t-SNE retains as much structural similarity information between data points in the original spaces as possible when reduced to a lower-dimensional space. The two axes represent the first and second dimensions of this new lower-dimensional space and the main purpose of these axes is to facilitate the visualization of high-dimensional relationships in a way that is intuitive and easy to interpret. As shown in Fig. 6, there are three clusters: one consisting of the majority of the samples, another consisting of some particle samples, and a third containing some unlabeled samples. This clustering occurs because the input data for these particle samples and unlabeled samples have similar characteristics. However, the data points of all classes are distributed very close together, often overlapping and forming a single cluster. This overlap poses a challenge for identifying and recognizing faults when using PD input signals.

TABLE 3. The hyper-parameter optimization.

Hyper-parameter	Minimum	Maximum	Type
Epochs	50	250	Integer
Number of layers	5	10	Integer
Kernel size	2 x 2	5 x 5	Integer
Learning rate	0.0001	0.001	Real
Weight decay	0.0001	0.001	Real
Dropout rate	0.1	0.8	Real
Random crop (c%)	10	40	Integer

Table 3 lists the ranges of each hyperparameter for hyperparameter optimization. Fig. 7 shows the observations regarding the trade-off between the quality and quantity of pseudo-labels, revealing that using a low value of the parameter for the confidence threshold cause a majority of predictions on unlabeled samples to exceed the confidence threshold. This collective influence significantly affects the unlabeled loss, as defined in Equation (11). By contrast, employing a higher value of the parameter for the confidence threshold value enables only a smaller fraction of higher-quality predictions on unlabeled samples to contribute to the unlabeled batch loss, resulting in a lower loss in the test set.

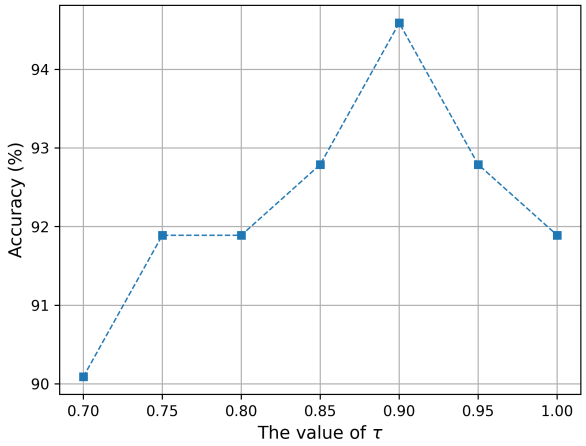

FIGURE 7. The parameter τ for confidence threshold.

For a performance comparison, the Self-SL [42], FixMatch [29] and baseline supervised learning (Base-SL) methods were implemented. Details of the model structures of the proposed SSL, Self-SL, FixMatch and Base-SL are listed in Table 4. For a fair and unbiased evaluation, consistency was maintained by adopting the same network architecture.

TABLE 4. Details of model structures.

Layer types	SSL	Self-SL	Base-SL	FixMatch
Input layer	3600 x 128	3600 x 128	3600 x 128	3600 x 128
Conv + ReLU	3 x 3, 8	3 x 3, 8	3 x 3, 8	3 x 3, 8
Max-pooling	4 x 1	4 x 1	4 x 1	4 x 1
Dropout	0.2	0.3	0.1	0.2
Conv + ReLU	3 x 3, 16	3 x 3, 16	3 x 3, 16	3 x 3, 16
Max-pooling	4 x 1	4 x 1	4 x 1	4 x 1
Conv + ReLU	3 x 3, 16	3 x 3, 16	3 x 3, 16	3 x 3, 16
Max-pooling	2 x 2	2 x 2	2 x 2	2 x 2
Conv + ReLU	3 x 3, 16	3 x 3, 16	3 x 3, 16	3 x 3, 16
Max-pooling	2 x 2	2 x 2	2 x 2	2 x 2
Conv + ReLU	3 x 3, 8	3 x 3, 8	3 x 3, 8	3 x 3, 8
Max-pooling	2 x 2	2 x 2	2 x 2	2 x 2
Flatten	3584	3584	3584	3584
Dropout	0.1	0.1	0.55	0.1
Dense + ReLU	1024	256	900	1024
Dropout	0.2	0.3	0.3	0.2
Dense layer	5	5	-	5
Softmax layer	-	-	5	-

Table 5 compares the classification performances of various models, including Base-SL(256) and Base-SL(624), which use 256 and 624 labeled samples for training,

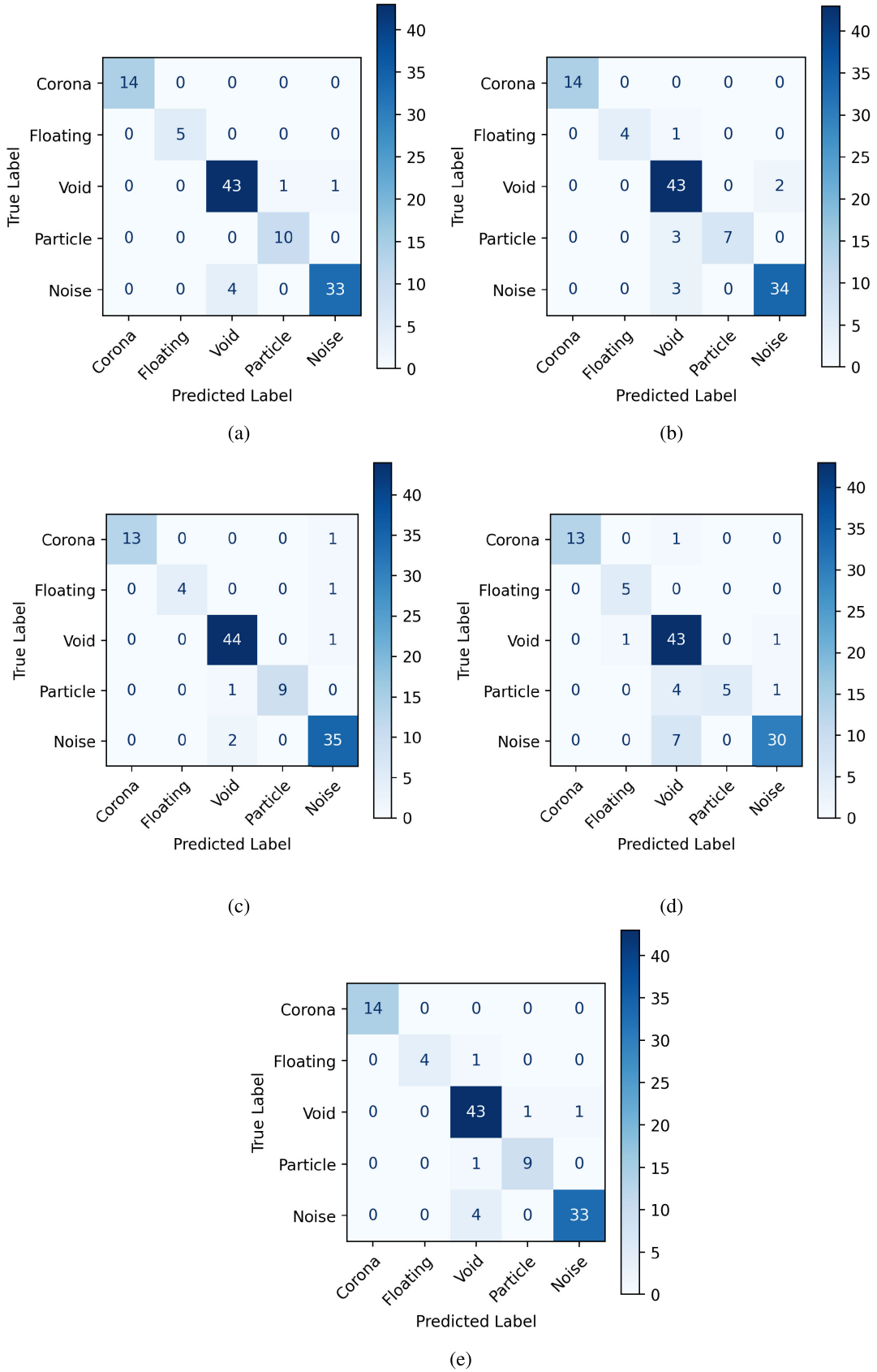

FIGURE 8. The confusion matrix of (a) Proposed SSL, (b) Base-SL(256), (c) Base-SL(624), (d) Self-SL, and (e) FixMatch.

TABLE 5. The accuracy comparisons.

Fault types	Overall	Corona	Floating	Void	Particle	Noise
Proposed SSL	**94.59%**	**100%**	**100%**	**95.56%**	**100%**	**89.19%**
Self-SL	92.79%	92.86%	100%	95.56%	50%	81.08%
Base-SL(624)	94.59%	92.86%	80%	97.78%	90%	94.59%
Base-SL(256)	91.89%	100%	80%	95.56%	70%	91.89%
FixMatch	92.79%	100%	80%	95.56%	90%	89.19%

TABLE 6. Precision, Recall and F1-score comparisons.

Metrics	Fault types	Overall	Corona	Floating	Particle	Void	Noise
Precision	**Proposed SSL**	**0.96**	**1**	**1**	**0.91**	**0.91**	**0.97**
	Self-SL	0.91	1	0.83	0.78	1	0.94
	Base-SL(624)	0.97	1	1	0.94	1	0.92
	Base-SL(256)	0.96	1	1	0.86	1	0.94
	FixMatch	0.95	1	1	0.88	0.90	0.97
Recall	**Proposed SSL**	**0.97**	**1**	**1**	**0.96**	**1**	**0.89**
	Self-SL	0.80	0.93	1	0.96	0.5	0.81
	Base-SL(624)	0.91	0.93	0.80	0.98	0.90	0.95
	Base-SL(256)	0.88	1	0.80	0.96	0.70	0.92
	FixMatch	0.91	1	0.80	0.96	0.90	0.89
F1-score	**Proposed SSL**	**096**	**1**	**1**	**0.93**	**0.95**	**0.93**
	Self-SL	0.85	0.96	0.91	0.86	0.67	0.87
	Base-SL(624)	0.94	0.96	0.89	0.96	0.95	0.93
	Base-SL(256)	0.91	1	0.89	0.91	0.82	0.93
	FixMatch	0.93	1	0.89	0.91	0.90	0.93

TABLE 7. Training and testing time comparisons.

Method	Training time (min)	Testing time (s)
Proposed SSL	**28.75**	$\sim$ **1**
Self-SL	38.67	$\sim$ 1
Base-SL(624)	12.67	$\sim$ 1
Base-SL(256)	9.33	$\sim$ 1
FixMatch	26.67	$\sim$ 1

TABLE 8. Performance comparisons of Base-SL methods using raw and augmented data.

Type	Accuracy (%)
Raw data (256)	85.59
Scale augmented	89.19
Crop augmented	87.39

respectively. The proposed SSL model, when trained with 256 labeled samples, shows an overall performance improvement of 2.7% compared to Base-SL(256) and 1.8% compared to FixMatch. Additionally, its accuracy surpasses Self-SL by 1.8% and matches the highest performance of Base-SL(624) at 94.59%. For specific fault categories such as corona, float, and particle, the proposed SSL method achieves exceptional accuracy, reaching 100%. In the case of void, its accuracy is only 2.22% lower than the peak accuracy of Base-SL(624), while for noise, it is 5.4% lower. The proposed SSL demonstrates higher classification performance in under-represented classes with small numbers of labeled data, such as corona, floating, and particle compared to over-represented classes with a large number of labeled data, including void and noise. This is because the proposed SSL with data augmentation improves the ability to learn feature representations in under-represented classes as the number of training data increases by using pseudo-labels. The confusion matrix shown in Fig 8 provides an overview of the performances of the methods in the classification problem.

In this study, because our data set was imbalanced, therefore *Precision*, *Recall* and $F1 - score$ were used as metrics to evaluate the classification performance and are respectively defined as

$$Precision = TP/(TP + FP), \tag{13}$$

$$Recall = TP/(TP + FN), \tag{14}$$

and

$$F1 - score = 2TP/(2TP + FP + FN) \tag{15}$$

where TP is the true positive, TN is the true negative, FP is the false positive and FN is the false negative. Table 6 presents Precision, Recall, and F1-score among Self-SL, FixMatch, Base-SL(624), Base-SL(256), and the proposed SSL. The proposed SSL exhibits notable precision and recall values, particularly achieving perfect scores of 1 in the corona and float classes. Its F1-score peaks at 1, surpassing other methods across most classes. Specifically, it achieves a precision of 0.91 for the particle class and a recall of 1 for the void class. The F1-score is the harmonic mean of the Precision and Recall, providing a balanced representation of both metrics in one metric. In conclusion, the proposed SSL method demonstrates superior overall performance with the highest F1-score of 0.96, surpassing Base-SL(256) and Base-SL(624) by 0.05 and 0.02, respectively. Additionally, the SSL model shows potential for further improvement with access to more unlabeled data, particularly through on-site

measurements. In contrast, Base-SL(624) faces challenges due to its reliance on additional labeled data.

Table 7 compares the training and testing time of the proposed SSL, Self-SL, FixMatch and Base-SL methods. Under the same hardware configuration, the proposed SSL had a training time shorter than that of Self-SL by approximately 10*min*. However, SSL was two and three times slower than Base-SL(624) and Base-SL(256), respectively. This was because the proposed SSL requires complex tasks such as pseudo-labeling, distribution alignment and curriculum learning for training. However, the proposed SSL had a testing time of 1*s*, with was not different from that of the other methods. Therefore, the proposed SSL can be used for PRPD fault diagnosis in GISs.

Table 8 presents a performance analysis of the Base-SL methods with the two data augmentation methods. Compared with Base-SL with raw data, Base-SL with scale augmented data showed a 3.6% performance improvement, whereas Base-SL with crop augmented data showed a 1.8% performance improvement. Therefore, data augmentation is a simple method that plays a major role in the success of the proposed SSL method, not only in increasing the size and diversity of the training set but also making our models more robust against noise.

V. CONCLUSION

In this study, we proposed a SSL method for fault diagnosis in GIS. The advantages of the proposed SSL include the utilization of pseudo-labeling for unlabeled PDs and integration of curriculum learning to optimize the total loss function. Experimental validation was conducted to assess the efficacy of the proposed SSL using on-site noise and PRPDs with fault types of corona, floating, void, and particle simulated using artificial cells. The experimental results showed that the proposed SSL had remarkable classification performance on limited labeled data and was superior to existing methods. In future studies, we intend to conduct PRPD experiments to obtain more measurements for fault data to verify the proposed method. In addition, the proposed SSL can be applied to the fault diagnosis of other power equipment.

REFERENCES

[1] G. Mazzanti, G. Stomeo, and S. Mancini, "State of the art in insulation of gas insulated substations: Main issues, achievements, and trends," *IEEE Elect. Insul. Mag.*, vol. 32, no. 5, pp. 18–31, Sep. 2016, doi: 10.1109/MEI.2016.7552373.

[2] A. Sabot, A. Petit, and J. P. Taillebois, "GIS insulation co-ordination: On-site tests and dielectric diagnostic techniques. A utility point of view," *IEEE Trans. Power Del.*, vol. 11, no. 3, pp. 1309–1316, Jul. 1996, doi: 10.1109/61.517485.

[3] B. Lu, W. Huang, J. Xiong, L. Song, Z. Zhang, and Q. Dong, "The study on a new method for detecting corona discharge in gas insulated switchgear," *IEEE Trans. Instrum. Meas.*, vol. 71, 2022, Art. no. 9000208, doi: 10.1109/TIM.2021.3129225.

[4] W. R. Si, J. H. Li, D. J. Li, J. G. Yang, and Y. M. Li, "Investigation of a comprehensive identification method used in acoustic detection system for GIS," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 17, no. 3, pp. 721–732, Jun. 2010, doi: 10.1109/TDEI.2010.5492244.

[5] X. Li, W. Liu, Y. Xu, and D. Ding, "Partial discharge and movement characteristics of micron-sized metal particles on insulator surface in gas-insulated switchgear with long-time AC stress," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 28, no. 6, pp. 2152–2160, Dec. 2021, doi: 10.1109/TDEI.2021.009775.

[6] M. Ren, M. Dong, Y. Liu, J. Miao, and A. Qiu, "Partial discharges in SF_6 gas filled void under standard oscillating lightning and switching impulses in uniform and non-uniform background fields," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 21, no. 1, pp. 138–148, Feb. 2014, doi: 10.1109/TDEI.2013.004114.

[7] S. Tao, F. Yang, D. Duan, W. Wang, X. Cheng, Z. Ren, H. Song, and J. Cui, "Transient Earth voltage detection technique for switchgears in distribution network," in *Proc. Int. Conf. Condition Monitor. Diagnosis (CMD)*, Sep. 2016, pp. 542–545.

[8] F. Zeng, J. Tang, X. Zhang, S. Zhou, and C. Pan, "Typical internal defects of gas-insulated switchgear and partial discharge characteristics," in *Simulation and Modelling of Electrical Insulation Weaknesses in Electrical Equipment*. London, U.K.: InTech, Oct. 2018, doi: 10.5772/intechopen.79090.

[9] S. Tenbohlen, S. Coenen, M. Djamali, A. Müller, M. Samimi, and M. Siegel, "Diagnostic measurements for power transformers," *Energies*, vol. 9, no. 5, p. 347, May 2016, doi: 10.3390/en9050347.

[10] S. Coenen and S. Tenbohlen, "Location of PD sources in power transformers by UHF and acoustic measurements," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 19, no. 6, pp. 1934–1940, Dec. 2012, doi: 10.1109/TDEI.2012.6396950.

[11] W. Gao, D. Zhao, D. Ding, S. Yao, Y. Zhao, and W. Liu, "Investigation of frequency characteristics of typical PD and the propagation properties in GIS," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 22, no. 3, pp. 1654–1662, Jun. 2015, doi: 10.1109/TDEI.2015.7116362.

[12] L. Li, J. Tang, and Y. Liu, "Partial discharge recognition in gas insulated switchgear based on multi-information fusion," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 22, no. 2, pp. 1080–1087, Apr. 2015, doi: 10.1109/TDEI.2015.7076809.

[13] L. Zhou, F. Yang, Y. Zhang, and S. Hou, "Feature extraction and classification of partial discharge signal in GIS based on Hilbert transform," in *Proc. Int. Conf. Inf. Control, Electr. Eng. Rail Transit (ICEERT)*, Lanzhou, China, Oct. 2021, pp. 208–213, doi: 10.1109/ICEERT53919.2021.00048.

[14] E. M. U. N. Ekanayake, H. M. N. P. Herath, G. C. Perera, J. V. Wijayakulasooriya, J. R. S. S. Kumara, M. A. R. M. Fernando, and K. M. K. S. Bandara, "Application of signal processing techniques for modeling of slot discharges in generators," in *Proc. 14th Conf. Ind. Inf. Syst. (ICIIS)*, Kandy, Sri Lanka, Dec. 2019, pp. 107–112, doi: 10.1109/ICIIS47346.2019.9063329.

[15] E. Ogliari, M. Sakwa, J. Wei, W. Liu, B. Schubert, and M. Palo, "General machine learning-based approach to pulse classification for separation of partial discharges and interference," *IEEE Sensors J.*, vol. 23, no. 21, pp. 26839–26849, Nov. 2023, doi: 10.1109/JSEN.2023.3317331.

[16] X. Zhang, S. Xiao, N. Shu, J. Tang, and W. Li, "GIS partial discharge pattern recognition based on the chaos theory," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 21, no. 2, pp. 783–790, Apr. 2014, doi: 10.1109/TDEI.2013.004020.

[17] N. K. Chauhan and K. Singh, "A review on conventional machine learning vs deep learning," in *Proc. Int. Conf. Comput., Power Commun. Technol. (GUCON)*, Greater Noida, India, Sep. 2018, pp. 347–352, doi: 10.1109/GUCON.2018.8675097.

[18] M.-T. Nguyen, V.-H. Nguyen, S.-J. Yun, and Y.-H. Kim, "Recurrent neural network for partial discharge diagnosis in gas-insulated switchgear," *Energies*, vol. 11, no. 5, p. 1202, May 2018.

[19] T.-D. Do, V.-N. Tuyet-Doan, Y.-S. Cho, J.-H. Sun, and Y.-H. Kim, "Convolutional-neural-network-based partial discharge diagnosis for power transformer using UHF sensor," *IEEE Access*, vol. 8, pp. 207377–207388, 2020, doi: 10.1109/ACCESS.2020.3038386.

[20] X. Zhou, X. Wu, P. Ding, X. Li, N. He, G. Zhang, and X. Zhang, "Research on transformer partial discharge UHF pattern recognition based on CNN-LSTM," *Energies*, vol. 13, no. 1, p. 61, Dec. 2019, doi: 10.3390/en13010061.

[21] J. Tang, M. Jin, F. Zeng, X. Zhang, and R. Huang, "Assessment of PD severity in gas-insulated switchgear with an SSAE," *IET Sci., Meas. Technol.*, vol. 11, no. 4, pp. 423–430, Jul. 2017.

[22] S. Salman and X. Liu, "Overfitting mechanism and avoidance in deep neural networks," 2019, *arXiv:1901.06566*.

[23] S. Albelwi, "Survey on self-supervised learning: Auxiliary pretext tasks and contrastive learning methods in imaging," *Entropy*, vol. 24, no. 4, p. 551, Apr. 2022, doi: 10.3390/e24040551.
[24] L. Schmarje, M. Santarossa, S.-M. Schröder, and R. Koch, "A survey on semi-, self- and unsupervised learning for image classification," *IEEE Access*, vol. 9, pp. 82146–82168, 2021, doi: 10.1109/ACCESS.2021.3084358.
[25] P. Arsomngern, C. Long, S. Suwajanakorn, and S. Nutanong, "Self-supervised deep metric learning for pointsets," in *Proc. IEEE 37th Int. Conf. Data Eng. (ICDE)*, Chania, Greece, Apr. 2021, pp. 2171–2176, doi: 10.1109/ICDE51399.2021.00219.
[26] A. Saha, A. Tejankar, S. A. Koohpayegani, and H. Pirsiavash, "Backdoor attacks on self-supervised learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 13327–13336.
[27] N. A. A. Braham, L. Mou, J. Chanussot, J. Mairal, and X. X. Zhu, "Self supervised learning for few shot hyperspectral image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Kuala Lumpur, Malaysia, Jul. 2022, pp. 267–270, doi: 10.1109/IGARSS46834.2022.9884494.
[28] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2013, vol. 3, no. 2, p. 896.
[29] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C. L. Li, "FixMatch: Simplifying semi-supervised learning with consistency and confidence," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 596–608.
[30] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. A. Raffel, "MixMatch: A holistic approach to semi-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.
[31] J. Jiang and W. Deng, "Boosting facial expression recognition by a semi-supervised progressive teacher," *IEEE Trans. Affect. Comput.*, vol. 14, no. 3, pp. 2402–2414, Sep. 2023, doi: 10.1109/TAFFC.2021.3131621.
[32] Z. Chen, L. Shen, D. Han, S. Lu, and G. Zhang, "Research on GIS corona discharge PRPD pattern based on optical detection," in *Proc. IEEE Sustain. Power Energy Conf. (iSPEC)*, Perth, WA, Australia, Dec. 2022, pp. 1–5, doi: 10.1109/iSPEC54162.2022.10033009.
[33] R. Piccin, A. R. Mor, P. Morshuis, A. Girodet, and J. Smit, "Partial discharge analysis of gas insulated systems at high voltage AC and DC," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 22, no. 1, pp. 218–228, Feb. 2015, doi: 10.1109/TDEI.2014.004711.
[34] D. Berthelot, N. Carlini, E. D. Cubuk, A. Kurakin, K. Sohn, H. Zhang, and C. Raffel, "ReMixMatch: Semi-supervised learning with distribution alignment and augmentation anchoring," 2019, *arXiv:1911.09785*.
[35] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. 34th Intl. Conf. Mach. Learn.*, 2017, pp. 1321–1330.
[36] S. Laine and T. Aila, "Temporal ensembling for semi-supervised learning," 2016, *arXiv:1610.02242*.
[37] P. Soviany, R. T. Ionescu, P. Rota, and N. Sebe, "Curriculum learning: A survey," *Int. J. Comput. Vis.*, vol. 130, no. 6, pp. 1526–1565, 2022.
[38] N.-Q. Dang, T.-T. Ho, T.-D. Vo-Nguyen, Y.-W. Youn, H.-S. Choi, and Y.-H. Kim, "Supervised contrastive learning for fault diagnosis based on phase-resolved partial discharge in gas-insulated switchgear," *Energies*, vol. 17, no. 1, p. 4, Dec. 2023, doi: 10.3390/en17010004.
[39] D. Berthelot, R. Roelofs, K. Sohn, N. Carlini, and A. Kurakin, "AdaMatch: A unified approach to semi-supervised learning and domain adaptation," 2021, *arXiv:2106.04732*.
[40] G. P. Kumar, G. S. Priya, M. Dileep, B. E. Raju, A. R. Shaik, and K. V. S. H. G. Sarman, "Image deconvolution using deep learning-based Adam optimizer," in *Proc. 6th Int. Conf. Electron., Commun. Aerosp. Technol.*, Coimbatore, India, Dec. 2022, pp. 901–904, doi: 10.1109/ICECA55336.2022.10009073.
[41] T. T. Cai and R. Ma, "Theoretical foundations of t-SNE for visualizing high-dimensional clustered data," *J. Mach. Learn. Res.*, vol. 23, no. 301, pp. 1–54, 2022.
[42] T. Chen, S. Kornblith, M. Norouzi, and G. E. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. 37th Int. Conf. Mach. Learn.*, vol. 119, 2020, pp. 1597–1607.

HO TRONG TAI received the bachelor's degree in control engineering and automation from the University of Science and Technology and the University of Da Nang (DUT), Da Nang, Vietnam, in November 2022. He is currently pursuing the master's degree with the Data Science Laboratory, Department of Computer Science and Information, Korea National University of Transportation (KNUT), South Korea. His research interests include artificial intelligence for smart grids, deep-learning applications for fault diagnosis, and signal processing in power systems.

YOUNG-WOO YOUN (Associate Member, IEEE) received the B.S. and M.S. degrees from Korea Advanced Institute of Science and Technology (KAIST), in 2005 and 2007, respectively, where he is currently pursuing the Ph.D. degree with the Kim Jaechul Graduate School of Artificial Intelligence. Since 2007, he has been a Researcher with Korea Electro-Technology Research Institute (KERI), Gyeongsangnam, South Korea. His research interests include condition monitoring, signal processing, and artificial intelligence for smart grid and high voltage power devices.

HYEON-SOO CHOI received the B.S. and M.S. degrees from Catholic Kwandong University, in 1998 and 2001, respectively. From 2001 to 2006, he was the Manager with Taekwang E&C, Seoul, South Korea. From 2006 to 2010, he was the Director with TEC&CO, Seoul. Since 2017, he has been the Director with Genad Systems, Jeollanam, South Korea. His research interests include condition monitoring, fault diagnosis, and machine learning.

YONG-HWA KIM (Member, IEEE) received the B.S. degree in electrical engineering and the Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 2001 and 2007, respectively. From 2007 to 2011, he was a Senior Researcher with Korea Electrotechnology Research Institute (KERI), Geonggi, South Korea. From 2011 to 2013, he was an Assistant Professor with the Division of Maritime Electronic and Communication Engineering, Mokpo National Maritime University, South Korea. From 2013 to 2021, he was a Professor with the Department of Electronic Engineering, Myongji University, South Korea. Since April 2021, he has been a Faculty Member with the Department of AI Data Engineering, Korea National University of Transportation, South Korea. His research interests include communication systems, digital signal processing, artificial intelligence for communications, radar systems, and the smart grids.