IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 14 February 2024, accepted 19 February 2024, date of publication 26 February 2024, date of current version 11 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3370241

RESEARCH ARTICLE

Shared Knowledge-Based Contrastive Federated Learning for Partial Discharge Diagnosis in Gas-Insulated Switchgear

VO-NGUYEN TUYET-DOAN[1], YOUNG-WOO YOUN[2,3], (Associate Member, IEEE), HYUN-SOO CHOI[4], AND YONG-HWA KIM[5], (Member, IEEE)

[1]Department of Electronic Engineering, Myongji University, Yongin-si, Gyeonggi-do 17058, Republic of Korea
[2]Smart Grid Research Division, Korea Electrotechnology Research Institute, Gwangju-si 61751, Republic of Korea
[3]Kim Jaechul Graduate School of Artificial Intelligence, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea
[4]Genad System, Naju-si, Jeollanam-do 58296, Republic of Korea
[5]Department of Artificial Intelligence, Korea National University of Transportation, Uiwang-si, Gyeonggi-do 16106, Republic of Korea

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

This work was supported in part by Korea Institute of Energy Technology Evaluation and Planning (KETEP) funded by Korean Government (MOTIE) under Grant 20225500000120, and in part by the National Research Foundation of Korea (NRF) funded by Korean Government (MSIT) under Grant 2022R1F1A1074975.

ABSTRACT Recently, deep neural networks have shown remarkable success in fault diagnosis in power systems using partial discharges (PDs), thereby enhancing grid asset safety and reliability. However, the prevailing approaches often adopt centralized large-scale datasets for training, without taking into account the impact of noise environments for Intelligent Electronic Devices (IEDs). Noise environments for PD measurements in gas-insulated switchgear (GIS) introduce variations in feature distributions and class representations, challenging the generalization ability of the trained models in new and diverse conditions. In this study, we propose a Shared Knowledge-based Contrastive Federated Learning (SK-CFL) for PD diagnosis in different noise environments for IEDs. The proposed SK-CFL combines federated learning principles with contrastive learning, empowering IEDs to collaboratively learn and share knowledge as regards PD and noise patterns. The proposed framework can learn representations between the same patterns across different IEDs while ensuring data privacy. Experimental results for PD diagnosis in GIS show that the proposed SK-CFL achieves a performance improvement in fault diagnosis, particularly in new and unseen environments. Specifically, the recall for unknown noise in untrained IED 6 demonstrates 92.86% of the proposed SK-CFL, in comparison with 64.29% and 35.71% of the conventional FL and baseline method, respectively. These results suggest that the proposed SK-CFL approach promises more adaptable, and resilient data-driven approaches since it protects data privacy that can operate effectively in challenging real-world environments.

INDEX TERMS Contrastive learning, deep neural networks, federated learning, gas-insulated switchgear, partial discharge.

I. INTRODUCTION

In recent decades, the global energy industry has experienced rapid growth owing to the development of the economy and society, resulting in a considerable increase in demand for electric energy. Gas Insulated Switchgear (GIS) is one of the most important components in terms of maintaining stable operational conditions of a power system, and its safety directly affects the reliability of the power grid [1], [2], [3], [4], [5], [6], [7], [8]. As time goes by, however, working under factual conditions affected by extreme operational, environmental, and weather conditions, insulation systems of electrical machines and power cables withstand increasingly

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

high levels of stress and contamination that will seriously threaten the stability of the power grid [1], [2], [3]. Partial discharge (PD) occurs when there is a partial breakdown of electrical discharge in GIS, which can happen near the high-voltage conductor or other locations [4], [5]. In GIS, partial discharge signals generally appear in the early stages of insulation deterioration, and multiple defects result in different forms of partial discharge signals during prolonged operations. Their causes in GIS vary, and different types can cause varying degrees of damage to GIS equipment [1], [6].

Partial discharge (PD) activity has been investigated using phase-resolved PD (PRPD) characteristics in GIS, where the amplitude of PD pulses is synchronized with the grid voltage applied. These characteristics facilitate the identification of the type of source present by analyzing the number of the PD pulses, and the maximum or average amplitude in each phase, as each type of PD (corona, superficial or internal) has a characteristic PRPD pattern [2], [7]. However, a major problem in PD identification is interferences of electrical noise, particularly in industrial environments where large noise sources or different sources of PD can act simultaneously, which may render the diagnosis and location of PD signals difficult, or even impossible. In several environments, trigger levels of noises were captured individually for each pattern based on the signal-to-noise ratio (SNR) of each sensor [8], [9], [10]. Additionally, if the noise trigger levels are extremely high, it may be challenging to detect other incipient PD sources that are associated with severe insulation failure [8].

Deep neural networks have emerged as one of the most active and powerful technologies widely used for diagnosing partial discharges (PD). By leveraging the advanced and mathematical evolution of machine learning, deep learning methods have harnessed the potential of stochastic optimization and multiple-layer structures of neural networks to improve the pattern recognition ability of PDs [11], [12].

Several deep neural network architectures have been explored to tackle the PD diagnosis challenge [13], [14], [15]. Convolutional Neural Networks (CNNs) are applied to recognize PD patterns using a dataset of 3500 samples and successful outcomes [13]. CNN architecture is utilized to identify six types of discharge defects, obtaining excellent classification results by leveraging the phase-amplitude response from a PD signal to diminish the input size [14]. A CNN-based Siamese structure is studied for PD diagnosis in Gas-Insulated Switchgear (GIS), even with limited data [15]. However, the conventional CNN-based PD classification methods used to develop the model via learning on the central server did not consider a data privacy issue. In the practical industrial scenario, there are different noise environments for IEDs. Further, the generalization of the trained models from one certain platform in real-world settings is difficult to assess due to the scarcity of available PD samples [16].

To overcome the shortcomings of central learning, federated learning methods have been globally adopted. They have been embraced by major companies as they play a crucial role in ensuring privacy-sensitive applications in scenarios where training data is collected from multiple clients [17], [18], [19], [20]. The success of federated learning is not limited to a particular domain but has extended its impact across several fields, achieving outstanding performances in diverse applications. In communication [21], federated learning has facilitated the development of robust and efficient communication systems. In mobile Internet technology [22], it has been leveraged to optimize user experience and enhance network performance. In natural disaster analysis [23], federated learning has demonstrated its capability to process and analyze large amounts of sensor data from different locations, enabling more accurate and timely disaster predictions and responses. In autonomous driving [24], federated learning has been utilized to train sophisticated models that enhance the safety and intelligence of self-driving vehicles. In heavy haul railway control [25], this approach has been successfully deployed to optimize train operations, improve scheduling efficiency, and ensure smoother rail transportation. Despite these achievements, federated learning algorithms rely solely on optimizing models using cross-entropy loss by aligning the marginal distribution without considering the fine-grained class distribution within each conditional environment, challenging fully capturing the diversity of real-world noise environments. The presence of diverse noise types can introduce variations in the feature distributions and class representations, challenging the generalization capability of the trained models when dealing with new and previously unseen environments.

To address limited federated learning algorithms, we propose a novel Shared Knowledge-based Contrastive Federated Learning (SK-CFL) for PD diagnosis in various noise environments. The proposed method leverages the power of federated learning, allowing individual IEDs to train locally using their own data. Further, the proposed SK-CFL integrates contrastive learning techniques, enabling IEDs to collaboratively learn and share knowledge, and promoting the discovery of shared patterns across different noise environments. Contrastive learning has the advantage of aggregating mutual knowledge between similar classes, even from new distributions in a way that brings similar samples (positive pairs) closer together while pushing dissimilar samples (negative pairs) apart. This has been successfully applied to various domains, including natural language processing, computer vision, and speech recognition [26], [27], [28], [29], [30], [31], [32].

The proposed SK-CFL addresses the challenge of enhancing model performance across all IEDs by leveraging knowledge learned locally. The proposed SK-CFL introduces a joint optimization approach by combining two loss functions: the Cross-Entropy loss (CE) and the class-supervised contrastive loss (CS). This combination aims to achieve

task-specific performance improvement within each IED and better generalization across different IEDs in the federated learning setting. The key contributions of SK-CFL are summarized as follows:

- The proposed SK-CFL allows IEDs to perform local training independently using their own datasets without data sharing. By utilizing local knowledge, each IED learns discriminative representations for the specific fault classification task within its environment. This approach ensures privacy protection and security risk while enhancing the model's capability to capture domain-specific patterns.
- The proposed SK-CFL incorporates contrastive learning (a powerful technique from visual representation learning), which enhances the model to learn mutual information on PD faults across different IEDs in GIS. Here, the model brings similar patterns closer and pushes dissimilar patterns apart.
- By jointly optimizing the CE and CS losses, the proposed SK-CFL aims to achieve improved fine-grained performance within each IED and generalization across different IEDs. This facilitates the development of robust and effective fault classification models that can be applied to new, unseen environments while benefiting from the diverse knowledge learned locally by each IED.
- Experimental results show that the proposed SK-CFL not only ensures outstanding PD diagnosis performances within specific IEDs but also enhances adaptability in new and unforeseen environments.

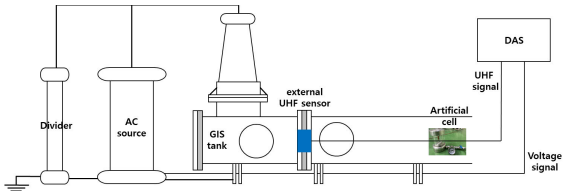
The remainder of this paper is organized as follows: Section II presents the description of experimental data for PRPDs and noise measurements in GIS. Section III discusses the problem formulation for the PD diagnosis and proposes SK-CFL by combining the federated learning framework with the contrastive learning techniques. Section IV presents the performance evaluations. Additionally, it compares the performance of the proposed method to other conventional methods in this section. Finally, Section V concludes the study.

II. EXPERIMENTAL DATA ANALYSIS

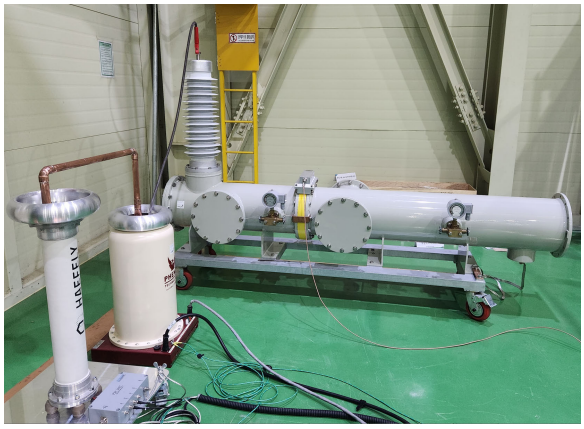
In this section, we present an experimental setup and the experimental results for PRPDs and on-site noise to investigate the assessment of PD characteristics using an ultra-high frequency (UHF) sensor in GIS.

A. EXPERIMENTAL SETUP

Fig. 1 illustrates an experimental system to detect PD signals in a GIS. The partial discharge experiment was conducted by applying AC voltages to artificial cells, which are installed in a GIS chamber. A cavity-backed patch antenna is used as an external UHF sensor and the signal from the UHF sensor is amplified with a gain of 45 dB which occurs within an operating bandwidth that ranges from 500 MHz to 1.5 GHz [33], [34], where the amplifier gain is set considering



(a)



(b)

FIGURE 1. An experimental system for the GIS: (a) Block diagram and (b) high-voltage test site.

the maximum amplitude of the partial discharge pulse. The amplified signal is then fed into a peak detector to capture the maximum value of each UHF PD pulse. The data acquisition system (DAS) further uses an analog-to-digital converter (ADC) with a sampling frequency of $1024 \times f_m$ samples per second, where $f_m = 60$ Hz is the power frequency, to digitize the signal. The maximum value of the digitized signal is captured at every 8 sample in the DAS, and 128 samples in each power cycle are applied for PRPD measurements. The PRPD measurements are obtained using an eight-bit ADC with a sampling rate of $P \times f_m$ samples per second, where $P = 128$ defines the number of data points in each power cycle and $fm = 60$ Hz is the power frequency. The measured signal comprises the pth data point and the mth power cycle is expressed as $x(m, p) \in \{0, 1, \ldots, 255\}$. This signal for $M = 3600$ power cycles is then formed in a matrix, X, where each element of the matrix represents the measured signal at a specific data point in a specific power cycle.

The measured signal for $M = 3600$ power cycles is denoted in form of matrix as

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

where $x(m, p) \in \{0, 1, \cdots, 255\}$ represents the measured signal with the value at the p-th data point for the m-th power cycle.

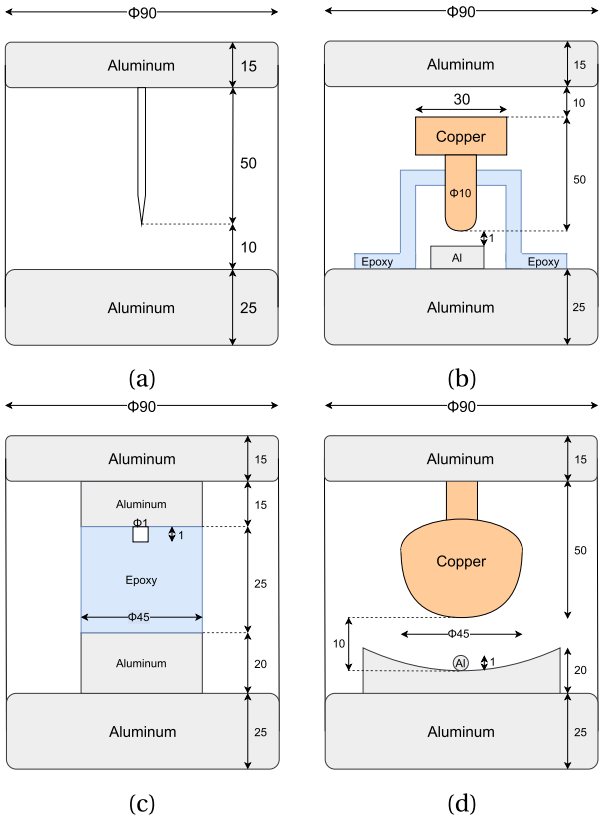


FIGURE 2. Artificial cells: (a) corona, (b) floating, (c) particle, and (d) void PDs.

B. PRPD MEASUREMENTS AND CHARACTERIZATIONS

For GIS, We investigated four types of faults (corona electrodes, floating electrodes, void defects, and free particles) using artificial cells [33], [34]. Each fault was simulated in a separate artificial cell (Fig. 2), as actual failures in GIS are rare in on-site environments. To simulate corona discharge, a sharp protrusion fixed on an electrode was used to create a local electric field enhancement through a needle with a 10 μm tip radius and a 1 mm in diameter (Ogura [35]), with a distance of 10 mm between the needle and the ground electrode, and a test voltage of 11 kV. To simulate an unconnected cell, a fabricated floating electrode cell was used with a test voltage of 10 kV. Here, small voids with a test voltage of 8 kV were formed between the epoxy disc and the upper electrode to simulate artificial void discharge. Free particle discharge was simulated using a small sphere with a 1 mm diameter placed on a concave ground electrode, with the HV electrode connected to a larger sphere with a diameter of 45 mm, fixed at a distance of 10 mm from the ground electrode and a test voltage of 10 kV. The artificial cells are made of aluminum on the top and bottom and acrylic on the sides. For stable experiments considering the material and thickness of the cells, the four artificial cells were filled with SF6 gas at 0.2 MPa. Experiments were conducted for each fault separately.

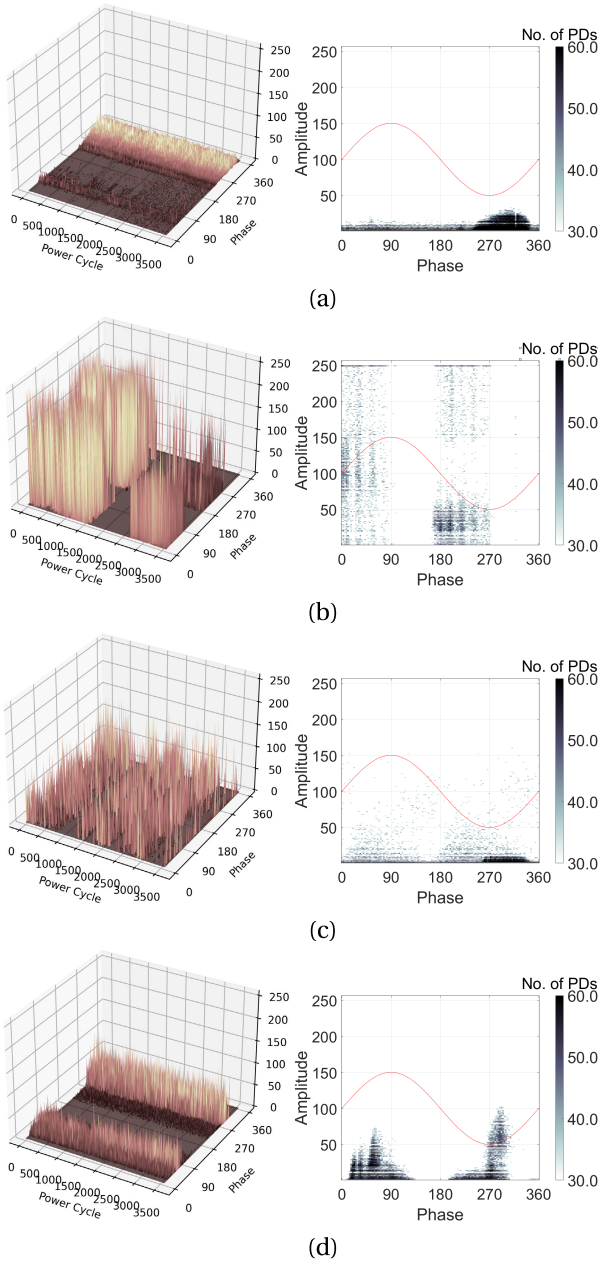


FIGURE 3. Sequential phase-resolved PDs (PRPDs) for four fault types in GIS: (a) corona, (b) floating, (c) particle, and (d) void.

Fig. 3 illustrates the time-domain data for PRPDs captured from UHF sensors in three-dimensional (3D) and two-dimensional (2D) representations. In Fig. 3, PRPDs are obtained from artificial cells for four types of insulation defects and include noises from the measurement system. Moreover, the 2D is performed by the 3D array: each PD event is represented by a pixel whose position corresponds with the phase angle and apparent charge at which it occurred. If multiple PD pulses occur at the same phase and magnitude, the color of the pixel changes to indicate the pulse count. In Fig. 3a, the discharge pulses of the corona fault [36] can be observed separately at both positive and negative half-cycles

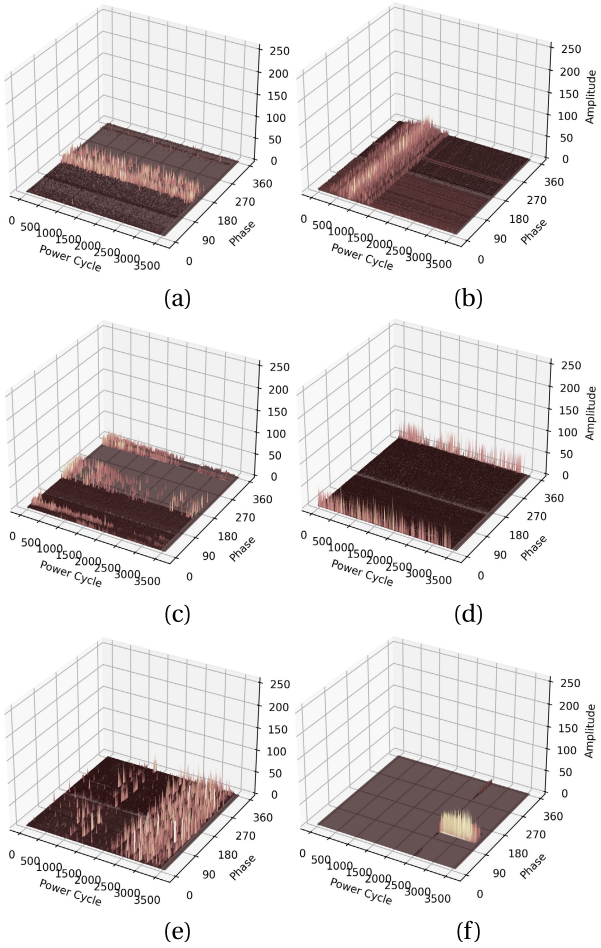


FIGURE 4. Noise measurements for (a) IED 1, (b) IED 2, (c) IED 3, (d) IED 4, (e) IED 5, and (f) IED 6.

with high frequencies ranging from 255 to 315 degrees, slightly from 45 to 90 degrees, and in the vicinity of zero, which have a similar distribution to that of void discharge. However, void discharge pulses exhibit a more sparse and high pattern. Also, discharge pulses for particle defects occur in the entire phase ranges, showing a difference in corona defects, as shown in Figs. 3a and 3c. For the floating and particle defects, there exists a very dense distribution of discharge pulses spanning various bands and intensity ranges, as shown in Figs. 3b and 3d, wherein the amplitude of the floating PDs reaches 250, close to the maximum measured value of 255 for the DAS.

C. NOISE MEASUREMENTS AND CHARACTERIZATIONS

The noise was measured using on-line UHF PD monitoring systems for on-site GISs in substations in South Korea. For noise measurements, diagnostic systems are installed on-site and noise is measured for GISs under normal conditions. The diagnostic systems use an 8-bit ADC to store raw data, and the amplitude of noise is quantified as an integer from 0 to 255. Fig. 4 shows on-site noise measurements collected by six different IEDs. Here, on-site noise signals from different

TABLE 1. Statistical analysis of noise levels.

	Noise 1	Noise 2	Noise 3	Noise 4	Noise 5	Noise 6
Min-of-Max	33	30	28	48	50	30
Max-of-Max	61	90	56	89	85	111
Means	0.104	0.828	0.206	0.282	0.190	0.062
Standard deviations	0.977	2.056	1.360	1.769	1.351	0.933

IEDs represent their own characteristic distribution depending on the location. Table 1 presents the statistical analysis for on-site noise measurements, where noise was classified into six types according to the measurement location. The statistical measures of interest were the minimum of maximum values, the maximum of maximum values, the mean, and the standard deviations, which were calculated separately for each noise level. The minimum and maximum values among the maximum values indicate the range of amplitude for each noise level. The mean values provide an estimate of the central tendency of the data for each noise level, while the standard deviations indicate the dispersion or variability of the data. These findings suggest that different noise levels may have distinct patterns or trends in their variability and central tendency.

A comparative statistical analysis of the amplitude levels of the different noises is illustrated in Fig. 5, where noise 1, noise 2, …, and noise 6 represent the noise measured at IED 1, IED 2, …, and IED 6, respectively. The almost maximum amplitude values of Noises 1, 2, and 3 are in the range of 28 to 61. However, there exists a notable difference in the distribution of amplitudes among them. The amplitudes of Noise 2 are concentrated around a mean value of 0.828, which is significantly higher than that of Noise 1, and Noise 3, with mean values of around 0.104, and 0.206, respectively. This indicates that Noise 2 has higher amplitude values compared to Noise 1 and Noise 3, despite having similar maximum amplitude values. Similar to Noise 1, Noise 6 also exhibits a mean amplitude value of approximately 0.933. However, the peak amplitude values for Noise 6 are located in [30,111] and have a wider range than Noise 1. Moreover, Noise 5 has a distribution with maximum values range of [50,85]. Despite having the lowest mean amplitude value among the analyzed noises, Noise 5 exhibits an equal standard deviation compared to Noise 1. In addition, Noise 4 displays a sparse distribution pattern with a range of mean values from [0.1,0.8] and a range of maximum values from [48,89].

III. PROPOSED SCHEME

In this section, we define the federated learning problem in PD diagnosis, before providing an overview of the proposed method's architecture for detecting PRPDs in a GIS. We propose a novel approach that combines the principles of federated learning and contrastive learning for PRPD fault diagnosis. The proposed Shared Knowledge-based Contrastive Federated Learning (SK-CFL) considers one server

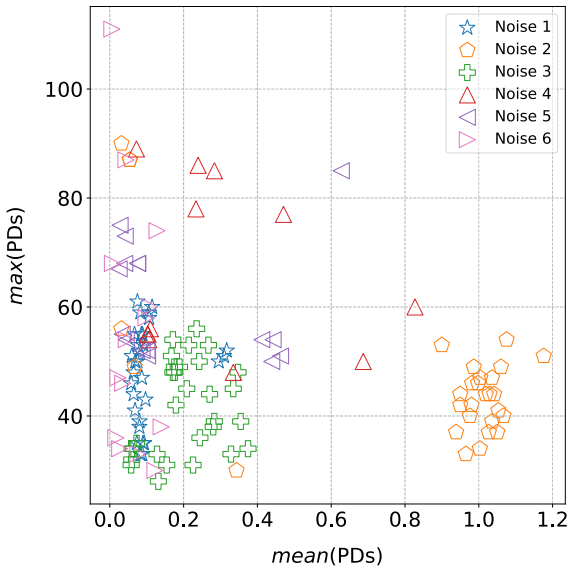


FIGURE 5. Visualizing statistical amplitude analysis of different noises.

and multiple intelligent electronic devices (IEDs). Federated Learning facilitates cooperative learning among individual environments, while Contrastive Learning refines feature representations, capturing both inter-class relationships and intra-class variations.

A. PROBLEM FORMULATION

Consider K IEDs $\mathcal{D}^1, \mathcal{D}^2, \cdots, \mathcal{D}^K$ with fault data under different noise conditions. Each IED $\mathcal{D}^k$ has its own local dataset comprising N^k samples represented as pairs $(\mathrm{X}_i^k, y_i^k)_{i=1}^{N^k}$, where X_i^k denotes the i-th sample and y_i^k is its corresponding label. In a conventional centralized learning scenario, all the data from different IEDs can be aggregated and trained by a global model. However, there are existing drawbacks to this centralized approach including concerns related to data privacy and security risks. Moreover, this approach can potentially hinder generalization owing to a lack of local knowledge, as IEDs often have domain-specific or location-specific knowledge that trains the model.

The proposed SK-CFL algorithm aims to improve the performance of the model across all the IEDs by combining the knowledge learned locally. Instead of transmitting the raw data, the proposed SK-CFL allows each IED to perform local model training independently using its data. Here, the central server aggregates the locally learned model parameters from each IED using federated learning, thereby updating the global model without direct access to the raw data. However, the simple optimization of the cross-entropy loss in the federated learning framework [37] without considering the cross-sample relationship can negatively impact the generalization performance in the presence of untrained target patterns, including new environment-collected or unseen patterns. To address this issue, we propose jointly optimizing our class-supervised contrastive loss. This approach considers the relationship between locals by maximizing the mutual information between similar classes across different locals while filtering task-irrelevant information caused by environmental changes. By so doing, our class-supervised contrastive loss captures shared class information and generates environment-independent class representations, thereby enhancing the generalization performance on untrained faults (faults from new environments, unseen faults).

The SK-CFL algorithm aims to further enrich the training process. By incorporating the class-supervised contrastive loss (L_{CS}) alongside the conventional Cross-Entropy loss (L_{CE}), SK-CFL endeavors to extract robust and discriminative feature representations. This joint optimization paradigm seeks to propel the generalization capabilities of the learned model across various IEDs and potentially unseen scenarios.

B. SHARED KNOWLEDGE-BASED CONTRASTIVE FEDERATED LEARNING (SK-CFL)

Fig. 6. shows the proposed SK-CFL architecture for fault classification. The proposed architecture involves multiple communication rounds between a central server and local IEDs. The key steps in the proposed SK-CFL process are as follows:

- **Step 1**: **Initialization**: The global model on the server is initialized with a pre-defined architecture with random weights and mutual representations.
- **Step 2**: **IED Training**: Each Intelligent Electronic Device (IED) independently performs local training using its own dataset. For each IED k, the local model parameters w_t^k are initialized with the global model's parameters w_{t-1}^G. Within each epoch, for each mini-batch b in the local dataset X^k and corresponding labels y^k:
  The local model is used to compute the cross-entropy loss l_{CE}.
  The b mini-batch samples concatenate with the mutual representations S, except its previous representations to calculate l_{CS}.
  The combined loss $Loss$ is calculated using a weighted combination of the cross-entropy loss (l_{CE}) and class-supervised contrastive loss (L_{CS}).
  The local model parameters are updated based on The $Loss$.
  Finally, randomly selected $I = 5$ samples for each class to aggregate feature representations $z_t^{k,C}$ and corresponding labels $y_t^{k,C}$.
- **Step 3**: **Update Model Parameters and Representations**: After completing the local training, each IED updates the model parameters w_t^k, local feature representations $z_t^{k,C}$, and corresponding labels $y_t^{k,C}$ to the central server. This allows the central server to incorporate the locally learned insights into the global parameters of the model.
- **Step 4**: **Server-side Aggregation**: On the server side, the aggregated global model parameters w_t^G are updated

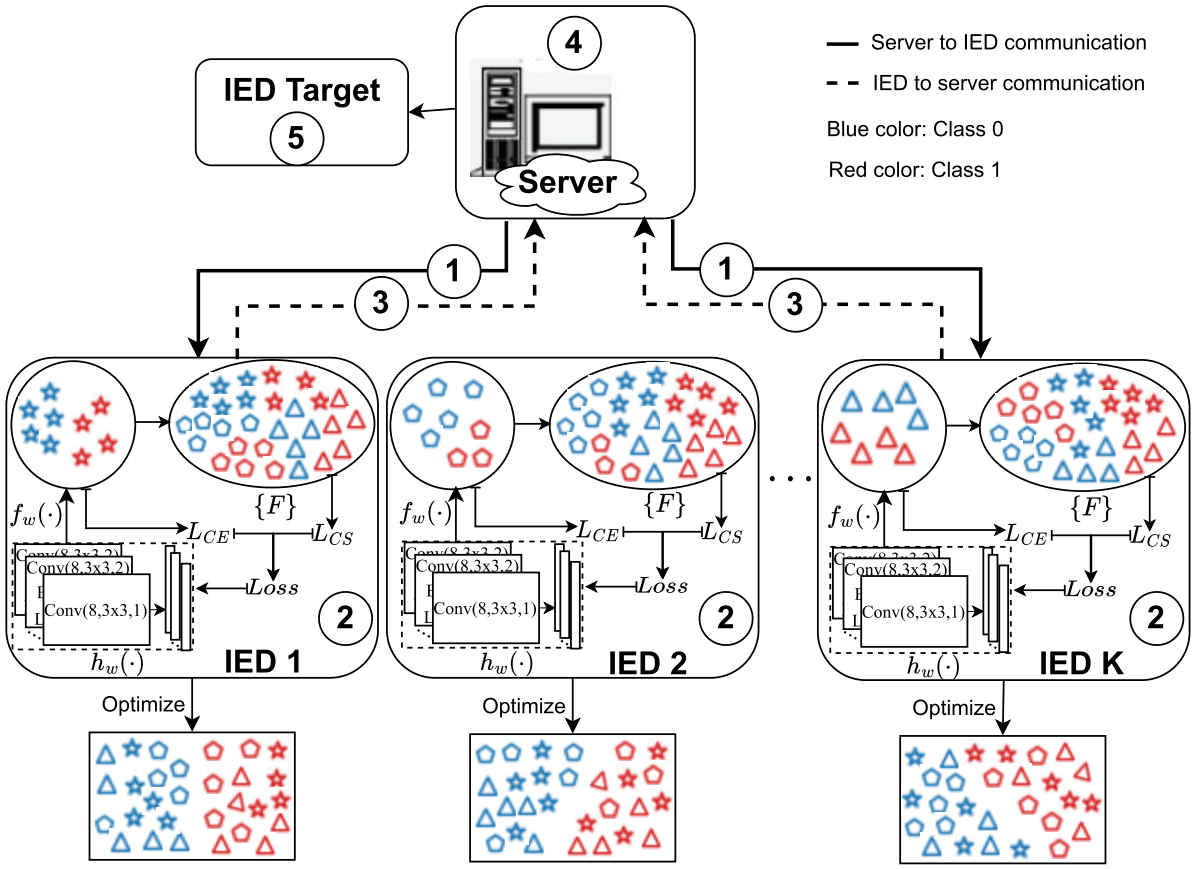


FIGURE 6. Proposed SK-CFL architecture.

through federated averaging. The server iterates through all the IEDs, retrieving their contributions to the global model. Additionally, the feature representations and labels (z_t^S and y_t^S) are aggregated from all the IEDs. The process from steps 2 to 4 is repeated for the number of rounds or until the global model converges and achieves the desired performance.

- **Step 5**: **Prediction for Untrained Targets**: After convergence, the updated global model W_{Global} is used to make predictions for new, unseen data (untrained IED targets).

The proposed architecture offers a collaborative approach to fault diagnosis. It leverages the power of federated learning to train the model on decentralized data sources while incorporating contrastive learning to improve the feature representations of the model. This combination makes it suitable for large-scale and distributed fault diagnosis systems, promoting efficient knowledge sharing among the IEDs. The overall proposed algorithm is described in Algorithm 1.

C. LOCAL TRAINING

The proposed SK-CFL method shares knowledge among the server and all the local IEDs to facilitate model aggregation of the network. The local model in the IEDs is the Convolutional Neural Network (CNN) due to its ability to automatically learn hierarchical representations from the raw input data in (1). The basic model architecture, as illustrated in Fig. 7, includes a feature extractor comprising a 6-layer CNN. Each of these convolutional layers employs 3×3-sized kernels and is equipped with 8 filters, coupled with a flattened layer before feeding to a Feedforward Neural Network (FNN) with 64 neurons. Further, a classifier is integrated into the architecture, comprising a 2-layer FNN with 64 neurons in each layer and an output layer for the probability of each class. Activation functions differ for these components. LeakyReLU activation functions are applied to the feature extractor, while the classifier utilizes ReLU activation functions.

During the training process, the parameters of the basis model are learned using mini-batches (denoted as $\mathcal{B}$) to minimize the following loss function, denoted as $J(w)$:

$$J(w) = \frac{1}{|\mathcal{B}|} \sum_{b \in \mathcal{B}} Loss(b), \tag{2}$$

where w represents the learnable parameters in the model. The loss function for each training sample in the mini-batch (b-th training sample in B) is calculated using the

Algorithm 1 Shared Knowledge-Based Contrastive Federated Learning (SK-CFL) Algorithm

Input: X^k, y^k, Number of rounds T, Number of epochs I, IED-list, Batch size B, Learning rate η, Temperature parameter τ, weighting factor α

IED Side Procedure:

def IEDTraining$(k, w_{t-1}^G, z_{t-1}^{S\setminus\{k\}}, y_{t-1}^{S\setminus\{k\}})$:
  Initialize $w_t^k \leftarrow w_{t-1}^G$
  for $n = 1$ to I epochs **do**
    for each mini-batch $b \in B$ **do**
      $l_{CE} \leftarrow h_{w_t^k}(X^{k,b}, y^{k,b})$
      $(z_t^{k,b}, y_t^{k,b}) \leftarrow f_{w_t^k}(X^{k,b}, y^{k,b})$
      $(z_t^F, y_t^F) \leftarrow$ (**Concat**$(z_t^{k,b}, z_{t-1}^{S\setminus\{k\}})$, **Concat**$(y_t^{k,b}, y_{t-1}^{S\setminus\{k\}})$
      $Loss \leftarrow (1-\alpha)\, l_{CE} + \alpha L_{CS}(z_t^F, y_t^F)$
      $w_t^k \leftarrow w_t^k - \eta \nabla Loss(w_{t-1}^{k\in b})$
    end for
    $z_t^{k,C} \leftarrow f_{w_t^k}(X^k) * 1_{k,\overline{1:I}\in C}$
    $y_t^{k,C} \leftarrow y^k * 1_{k,\overline{1:I}\in C}$
  end for
  return $w_t^k, z_t^{k,C}, y_t^{k,C}$

Server Side Procedure:

Initialize w_0^G, z_0^S, y_0^S
for $t = 1$ to T rounds **do**
  for each IED $k \in$ IED-list **do**
    $w_t^k, z_t^{k,C}, y_t^{k,C} \leftarrow$ **IEDTraining**$(k, w_{t-1}^G, z_{t-1}^{S\setminus\{k\}}, y_{t-1}^{S\setminus\{k\}})$
    $w_t^G \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_t^k$
    $(z_t^S, y_t^S) \cup_{k\in K} (z_t^{k,C}, y_t^{k,C})$
  end for
end for
Global model: $W_{Global} \leftarrow w_t^G$

Output: Global model W_{Global}.

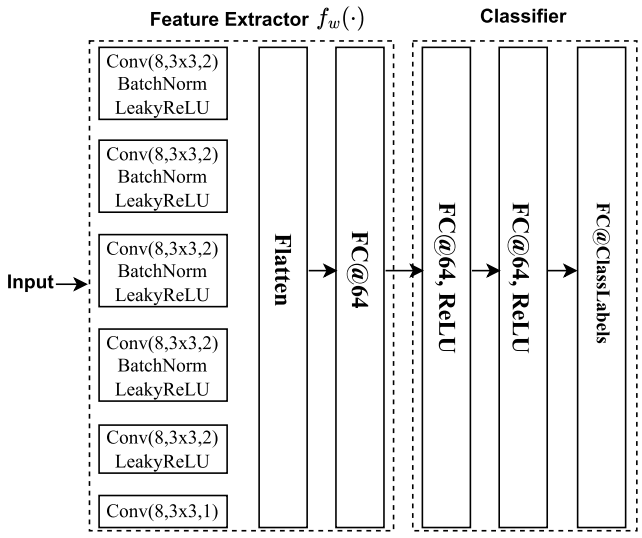


FIGURE 7. Basic architecture $h_w(\cdot)$ in the proposed SK-CFL framework.

cross-entropy loss as follows:

$$L_{CE} = -\sum_{i=1}^{C} e_i \log\left(z_i^{(b)}\right), \tag{3}$$

where C is the number of classes or categories in the dataset. The superscript b denotes the index for the b-th training sample within the mini-batch B. The term e_i is the value of 1 when the index i corresponds with the ground truth class for the sample and 0 otherwise. The Adam optimizer is adopted to update the learnable parameters of the network during the training process.

D. CONTRASTIVE LEARNING BASED MUTUAL INFORMATION CONNECTIONS

Contrastive learning constitutes a robust technique utilized to amplify the efficacy of vision-related tasks. This method hinges on the fundamental notion of discerning differences between pairs of samples to stimulate feature learning. It employs a neural network model denoted as $h_w(\cdot)$ to learn both common attributes shared among different data classes and attributes that distinguish one data class from another. The model $h_w(\cdot)$ operates by applying an encoder $f_w(\cdot)$ with parameters w to extract key feature representations z_l from input sample X_l, computed as $z_l = f_w(X_l)$, and a classifier for class probability estimation. In the training phase, each input batch of a dataset concatenates with the representations in set S, except its previous representations which is received from the server, $(z_t^F, y_t^F) \leftarrow$ (**Concat**$(z_t^{k,b}, z_{t-1}^{S\setminus\{k\}})$, **Concat**$(y_t^{k,b}, y_{t-1}^{S\setminus\{k\}})$ before considering an arbitrary pattern indexed as $i \in F$. This pattern corresponds with a sample pair denoted as X_i, y_i. The principal objective is to maximize the mutual information between X_i and samples from similar labels $y_i = y_p$, also referred to as positives, while minimizing samples from different labels $y_i \neq y_a$, called negatives. In the sub-contrastive learning algorithm, the loss of the sampled sample $\{X_i, y_i\}$ is mathematically expressed as [38]

$$L_{CS,i} = \frac{-1}{|P(i)|} \sum_{p\in P(i)} \log \frac{h_w(\{X_i, X_p\})}{\sum_{a\in A(i)} h_w(\{X_i, X_a\})} \tag{4}$$

$$= \frac{-1}{|P(i)|} \sum_{p\in P(i)} \log \frac{\exp(z_i \bullet z_p / \tau)}{\sum_{a\in A(i)} \exp(z_i \bullet z_a / \tau)}, \tag{5}$$

where $|P(i)|$ is the cardinality of all the positives, while $P(i) \equiv \{p \in A(i) : y_i = y_p\}$ is the set of indices of positive samples, known to belong to a similar label with X_i.

To maximize the mutual information between X_i and all the positive samples, a lower bound of mutual information is inferred by the supervised contrastive loss function throughout the process to optimize the model h_w and the encoder z. The optimal model h_w^{opt} is directly related to the density ratio between the joint distribution $p(z_i, z_p)$ and the product of marginals $p(z_i)p(z_p)$ [39], defined as

$$h_w^{opt}(\{X_i, X_p\}) \propto \frac{p(z_i, z_p)}{p(z_i)p(z_p)} \propto \frac{p(z_i|z_p)}{p(z_i)}, \tag{6}$$

where $\propto$ stands for "proportional to". By inserting this relationship into Equation (4), we can deduce that

$$
\begin{aligned}
L_{CS,i}^{opt} &= \frac{-1}{|P(i)|}\sum_{p\in P(i)}\log\left[\frac{\frac{p(z_i,z_p)}{p(z_i)p(z_p)}}{\sum_{a\in A(i)}\frac{p(z_i,z_a)}{p(z_i)p(z_a)}}\right] \\
&= \frac{1}{|P(i)|}\sum_{p\in P(i)}\log\left[\frac{\sum_{a\in A(i)}\frac{p(z_i,z_a)}{p(z_i)p(z_a)}}{\frac{p(z_i,z_p)}{p(z_i)p(z_p)}}\right] \\
&= \frac{1}{|P(i)|}\sum_{p\in P(i)} \\
&\quad \log\left[\frac{\frac{p(z_i,z_a)}{p(z_i)p(z_a)}\Big|_{a=p} + \sum_{a\in A(i)\setminus\{p\}}\frac{p(z_i,z_a)}{p(z_i)p(z_a)}}{\frac{p(z_i,z_p)}{p(z_i)p(z_p)}}\right] \\
&= \frac{1}{|P(i)|}\sum_{p\in P(i)} \\
&\quad \log\left[1+\frac{p(z_i)p(z_p)}{p(z_i,z_p)}\sum_{a\in A(i)\setminus\{p\}}\frac{p(z_i,z_a)}{p(z_i)p(z_a)}\right] \qquad (7) \\
&\approx \frac{1}{|P(i)|}\sum_{p\in P(i)}\log\left[1+\frac{p(z_i)p(z_p)}{p(z_i,z_p)}(|F|-1)\right] \\
&\geq \frac{1}{|P(i)|}\sum_{p\in P(i)}\log\left[\frac{p(z_i)p(z_p)}{p(z_i,z_p)}|F|\right] \\
&= \log(|F|)+\frac{1}{|P(i)|}\sum_{p\in P(i)}\log\left[\frac{p(z_i)p(z_p)}{p(z_i,z_p)}\right] \\
&= \log(|F|)-\sum_{p\in P(i)}\frac{1}{|P(i)|}\log\left[\frac{p(z_i,z_p)}{p(z_i)p(z_p)}\right] \\
&= \log(|F|)-\sum_{p\in P(i)} I(z_i;z_p). \qquad (8)
\end{aligned}
$$

Since $a \in A(i)\setminus\{p\}$ represents the number of negative samples or pairs $|F|-1$ in the training set of X_i. Eq. (8) becomes more precise as the number of samples $|F|$ increases. This prompts an inequality, $\sum_{p\in P(i)} I(z_i;z_p) \geq \log(|F|) - L_{cons,i}^{opt}$. This holds for every encoder z [39]. Therefore, this inequality underscores that by minimizing the $L_{CS,i}^{opt}$ objective function, the lower bound on the mutual information $\sum_{p\in P(i)} I(z_i;z_p)$ can be effectively maximized [39].

E. OVERALL LOSS

In the proposed method, we introduce a joint optimization approach that combines two objectives: the class-supervised contrastive loss (L_{CS}) and the cross-entropy loss (L_{CE}). This combined optimization aims to enhance the generalization performance of the model on new, unseen IEDs.

The class-supervised contrastive loss (L_{CS}) addresses the relationships between different noise environments by generalizing similar class representations. It encourages the model to learn compact and discriminative feature representations across different IEDs. The class-supervised contrastive loss is defined as follows:

$$L_{CS} = \frac{1}{n_F}\sum_{i=1}^{F} L_{CS,i}, \qquad (9)$$

where $F \in \{b \cup S\setminus\{k\}\}$ merges the b mini-batch samples and S samples in which it takes $I = 5$ samples for each class in K IEDs, except its k previous representations.

The overall loss for the classifier task is a linear combination of the cross-entropy loss and the class-supervised contrastive loss, which is controlled by the weighting coefficient α. The overall loss is formulated as follows:

$$Loss = (1-\alpha)\,L_{CE} + \alpha L_{CS}, \qquad (10)$$

where α represents the weighting coefficient that determines the relative importance of each loss term. By jointly optimizing the cross-entropy loss and class-supervised contrastive loss, the proposed method aims to achieve both task-specific performance improvement and better generalization across different IEDs in the federated learning setting.

IV. PERFORMANCE EVALUATION

This section presents the experimental results and performance evaluation using PRPD data in different noise environments.

A. EXPERIMENT RESULTS

Table 2 shows the number of experimental dataset for PRPDs and on-site noises in GIS, where four types of PRPDs such as corona, floating, particle, and void PDs, are considered. Noises were measured using the UHF sensor in on-site fields. Here, the same PRPD data from all IEDs and noise measured from each IED are used. Each PRPD fault and noise signal has 3600 power cycles (M=3600), where each power cycle has 128 data points (P=128). We partitioned the dataset into separate training and testing sets, following an 80:20 ratio for each fault category to monitor the model convergence on local IEDs during the training phase. For hyperparameter optimization, we conducted an exhaustive series of experiments to fine-tune the hyperparameters of our model (layer count, kernel size, and the number of kernels) and training algorithm configurations (learning rate, batch size, and epochs).

Our experimentation process resulted in an optimized configuration, as shown in Fig. 7. The feature extractor comprises a 6-layer CNN, utilizing 3×3-sized kernels and 8 filters in each layer. This is followed by a flattened layer before input to a Feedforward Neural Network (FNN) comprising 64 neurons. After that, a classifier consists of a 2-layer FNN with 64 neurons per layer and an output layer for class probability estimation. LeakyReLU activation functions were employed for the feature extractor, while ReLU activation functions were used for the classifier.

TABLE 2. Number of experimental datasets for the PRPDs and noises.

Local IED	CORONA	FLOATING	PARTICLE	VOID	Noise	Total
1	94	35	66	242	30	467
2	94	35	66	242	30	467
3	94	35	66	242	30	467
4	94	35	66	242	13	450
5	94	35	66	242	16	453
6	94	35	66	242	14	451

TABLE 3. Maximum and minimum bound of hyperparameter optimization.

Hyperparameter	Minimum	Maximum	Type
Number of layers	3	10	Integer
Kernel size	3×3	7×7	Integer
Stride size	1	3	Integer
Learning rate	0.0001	0.1	Real
Minibatch size	8	128	Integer
Temperature τ	0.01	1	Real
Contrastive weight α	0.01	1	Real
Number of epochs	1	5	Integer
Number of rounds	50	200	Integer

Additionally, batch normalization was incorporated into the feature extractor to enhance training stability. Detailed ranges for our approach can be found in Table 3. The reason for our choice of hyperparameters included the batch size, learning rate, temperature τ, contrastive weight α, number of epochs, and number of rounds of 32, 0.001, 0.7, 0.1, 2, 100, respectively. This configured setup resulted in the highest overall classification accuracy when applied to the trained PD data. To mitigate the impact of random initial values on the network, we conducted 5 trials in our experiments. After that, we averaged the results to affirm the robustness of the proposed model.

B. COMPARED MODELS

For comparison purposes, simulations are conducted based on three key aspects: Comparison of the model training efficiency between SK-CFL-Based versus Traditional Centralized Model Training. Comparison between the accuracy of the baseline method and FL model. Comparison of the model training efficiency between FedAvg vs. Proposed SK-CSL Method.

- Centralized Training (No Data Privacy): This method trains a centralized model using all IED's data without any data privacy restrictions. The results are compared two methods, including a Centralized Baseline method (Baseline-CT) [40] where the models are updated solely by the cross-entropy loss function and Centralized Contrastive learning method (CL-CT) [41] using both the cross-entropy loss and contrastive loss to optimize the model.
- Baseline method (Single dataset): In this case, the diagnosis model is trained on a dataset from only one IED and tested on the other untrained IEDs. For example, to assess the performance of an untrained IED 1, four separate diagnosis models are trained using their respective datasets (IED 2, IED 3, IED 4, and IED 5). These models are then individually tested with the data from IED 1, resulting in four distinct diagnosis accuracies. The final evaluation for the performance of untrained IED 1 is based on the average accuracy across these four tests.
- Federated Learning (FL) [37]: This represents the traditional Federated Averaging algorithm, where the models are trained on local IED datasets, and their parameters are averaged to create a global model. Here, all clients have equal weight during the training process.

To ensure a fair and unbiased evaluation, we maintained consistency by adopting the same loss function and network architecture for all baseline methods included in the comparison. In addition, all the experiments were conducted on the same PC executed on an Intel(R) UHD Graphics 770 GPU using PyTorch 1.7.

C. EXPERIMENTAL RESULTS

Table 4 presents the performance of fault diagnosis of training processes using different training methods. It includes central training methods (Baseline-CT and CL-CT) and local training methods (Baseline, Conventional FL, and Proposed SK-CFL) along with their training and testing accuracy scores. This is used to test the generalization performance of the local training IEDs by calculating the average accuracy on both the training and testing datasets. Notably, the testing dataset is not used for training or other local IED models. As can be seen from the table, the proposed SK-CFL method consistently demonstrates competitive accuracy in the training and testing across all untrained IED cases. Experimental results show that the proposed method achieves high accuracy not only during training but also in their own testing dataset even when the updated weights of IEDs by aggregated in the sever after each round.

Table 5 shows the results of the fault diagnosis from untrained IED datasets. From the perspective of central training algorithms, the CL-CT achieves higher accuracy in fault diagnosis compared to the Baseline-CT. This suggests that the introduction of contrastive learning techniques has a positive impact on the fault diagnosis by encouraging similar representations for similar data points, likely helping the model to better understand the underlying untrained patterns. From the perspective of local training methods, it can be inferred that FL-based schemes outperformed the baseline model significantly, indicating that FL-based methods are effective in addressing the challenge of generalization in untrained IEDs. FL enables individual IEDs to locally and then collaboratively refine a global model, resulting in better generalization. Moreover, the proposed SK-CFL method

TABLE 4. Performance (%) of fault diagnosis in the GIS of the training process on trained IED's dataset using different methods.

	Centralized training				Local training					
	Baseline-CT		CL-CT		Baseline		Conventional FL		Proposed SK-CFL	
Untrained	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
IED 1	98.40	97.85	98.13	97.85	99.53	95.72	98.52	97.65	98.42	96.71
IED 2	99.81	98.74	99.63	98.71	99.42	96.48	99.02	97.73	99.15	98.80
IED 3	99.60	99.09	99.35	98.98	99.94	96.05	98.27	97.37	99.26	98.47
IED 4	99.55	98.36	99.49	99.02	99.92	94.99	98.37	97.48	98.75	98.58
IED 5	99.33	98.66	99.09	98.26	99.97	95.10	98.12	96.51	99.46	98.32
IED 6	99.92	99.78	99.15	99.20	99.89	96.72	98.38	96.65	99.18	99.05
Average	99.44	98.75	99.14	98.84	99.78	95.84	98.38	97.23	99.03	99.04

TABLE 5. Accuracy (%) of fault diagnosis on the untrained IED dataset using different methods.

	Centralized training		Local training		
Untrained	Baseline-CT	CL-CT	Baseline	Conventional FL	Proposed SK-CFL
IED 1	98.75 ± 0.01	99.01 ± 0.01	96.19 ± 0.02	98.46 ± 0.02	99.31 ± 0.007
IED 2	98.32 ± 0.01	98.16 ± 0.01	95.52 ± 0.02	98.45 ± 0.006	98.76 ± 0.007
IED 3	99.14 ± 0.006	99.42 ± 0.005	95.41 ± 0.01	98.14 ± 0.006	98.75 ± 0.002
IED 4	98.97 ± 0.01	99.28 ± 0.002	97.84 ± 0.02	98.82 ± 0.004	99.61 ± 0.005
IED 5	98.75 ± 0.004	97.35 ± 0.02	95.47 ± 0.02	98.43 ± 0.004	99.15 ± 0.008
IED 6	97.02 ± 0.01	98.49 ± 0.003	95.35 ± 0.01	98.36 ± 0.007	98.53 ± 0.007
Average	98.49 ± 0.008	98.62 ± 0.008	95.96 ± 0.017	98.44 ± 0.018	99.02 ± 0.006

demonstrates the highest accuracy across all the cases. These results show the effectiveness of the proposed SK-CFL by combining federated and contrastive learning techniques to enable IEDs to learn collaboratively and share knowledge.

The comparison of the performance for the methods for different scenarios by combining randomly the untrained IEDs as targets for test performance is shown in Table 6. In this experiment, the accuracy of the approaches using local training is performed. It can be observed that the proposed SK-CFL demonstrates performance with an average accuracy of 98.21%, compared to the conventional FL and Baseline model with 97.36% and 96.06%, respectively. This indicates the robustness and superiority of the proposed SK-CFL in the generalization of untrained IEDs even in more complicated diagnostic tasks.

To highlight the advantages of the proposed SK-CFL technique, particularly in scenarios with significant noise variability owing to data collected from diverse noise environments, Fig. 8 illustrates a comparison of confusion matrices for the proposed SK-CFL, conventional FL, and Baseline method (which only uses IDE 3 for model training). The accuracy is tested on the untrained IDE 6. As shown in Fig. 8, the proposed SK-CFL algorithm achieves an accuracy of 92.86% with relatively high discrimination between the noise and fault types. In contrast, the conventional FL and Baseline model achieve accuracies of only 64.29%, and 35.71%, respectively, which is lower than 28.57%, and 57.15% than our proposed method, respectively. In particular, the proposed SK-CFL achieves higher accuracy for noise patterns on the untrained IED 6 compared to other methods. This is because the significant differences are observed in the distribution of noise patterns measured across different IEDs.

Table 7 presents the results of a 5-fold cross-validation comparing the performance of the conventional FL and proposed SK-CFL method in the scenario 3 and 4. The results of a paired t-test with a 95 percent confidence level to assess the significant difference of performances of two methods. As shown in table 7, the p-values obtained from the paired t-tests achieve both less than the significance level of 0.05 and these results are considered as ''significantly different'' because the p-value is less than 0.05 [42]. Therefore, there is a significant difference between the proposed SK-CFL and the traditional FL. On average, the accuracy for the proposed SK-FLC model is 1.18% higher than that of the traditional FL in the scenario 4.

Furthermore, we analyzed of the internal representations of the trained network at the output of the encoder. Figs. 9a, 9b, and 9c depict the t-distributed stochastic neighbor embedding (t-SNE) illustrations of the input data vectors from the untrained IDE 6 of the proposed SK-CFL, Conventional FL, and Baseline (using IDE 3 for training) schemes, respectively. The t-SNE is used to project high-dimensional vectors

TABLE 6. Accuracy (%) of the combinations for different untrained IEDs.

Scenario	Untrained IEDs	Baseline	Conventional FL	Proposed SK-CFL
S1	5, 6	95.41 ± 0.02	95.77 ± 0.02	97.26 ± 0.02
S2	1, 2	95.80 ± 0.002	96.72 ± 0.01	98.13 ± 0.01
S3	3, 4	96.63 ± 0.02	99.20 ± 0.002	99.44 ± 0.004
S4	4, 5, 6	96.22 ± 0.03	98.01 ± 0.005	98.14 ± 0.01
S5	3, 4, 5	96.24 ± 0.02	97.10 ± 0.008	98.09 ± 0.02
Average		96.06 ± 0.02	97.36 ± 0.009	98.21 ± 0.013

TABLE 7. Comparison of accuracy and t-test performance between the conventional FL and proposed SK-CFL method using a 5-fold cross-validation in scenarios 3 and 4.

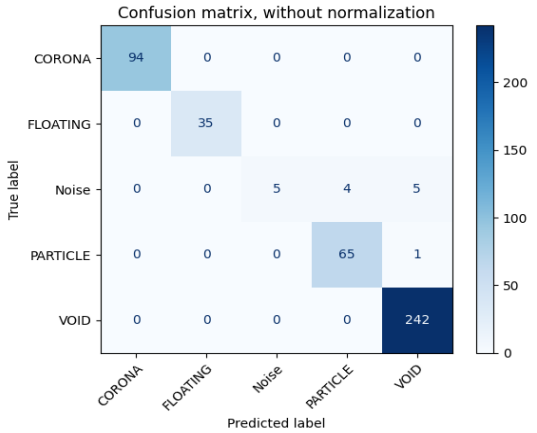
Scenario	K-Fold Cross Validation	Accuracy of Conventional FL	Accuracy of Proposed SK-CFL	Paired t-test p-value
S3	Fold 1	99.23	99.35	0.04
	Fold 2	98.58	99.35	
	Fold 3	99.56	99.56	
	Fold 4	98.59	99.35	
	Fold 5	98.80	99.67	
Average		**98.95**	**99.46**	
S4	Fold 1	97.42	97.79	0.005
	Fold 2	96.82	97.19	
	Fold 3	96.38	98.37	
	Fold 4	97.05	98.38	
	Fold 5	96.16	98.01	
Average		**96.77**	**97.95**	

TABLE 8. An analysis of precision, recall, and F-1 score for different classes on the untrained IDE 6.

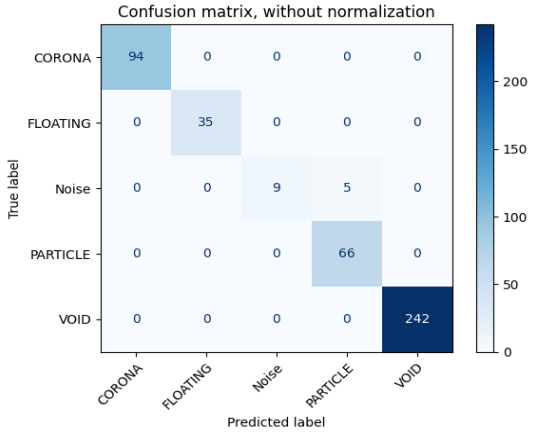
		Corona	Floating	Noise	Particle	Void
Precision	Baseline	100%	100%	100%	92.86%	100%
	Conventional FL	100%	100%	100%	92.96%	100%
	Proposed SK-CFL	100%	100%	100 %	**98.51%**	100%
Recall	Baseline	100%	100%	35.71%	98.48%	100%
	Conventional FL	100%	100%	64.29%	100%	100%
	Proposed SK-CFL	100%	100%	**92.86%**	100%	100%
F-1 score	Baseline	100%	100%	52.63%	95.59%	100%
	Conventional FL	100%	100%	78.26 %	96.35%	100%
	Proposed SK-CFL	100%	100%	**96.30%**	**99.25%**	100%

into 2D spaces while preserving pairwise similarity [44]. As shown in Fig. 9a, it is difficult to identify and recognize discharge faults precisely, especially for types of noise, particle, and void patterns. In Fig. 9b, the confusion between noise and particle fault can still happen. Contrarily, Fig. 9c reveals distinct clustering among all the classes. Hence, the proposed SK-CFL can learn fault class representations that are adaptable to new and unforeseen faults in a variety of noise environments.

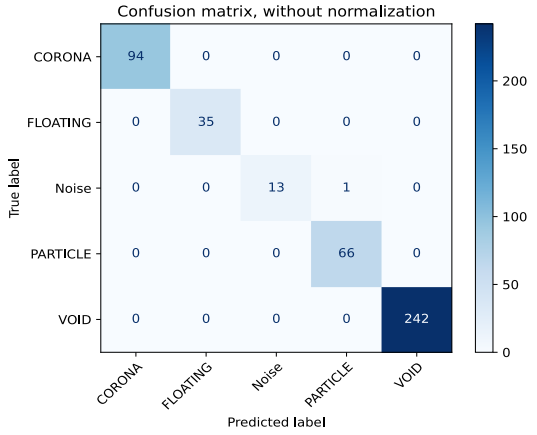
Table 8 presents an analysis of performance metrics of precision, recall, and F-1 score [43] for different classes from the untrained IDE 6. As can be seen from table 8, the proposed SK-CFL demonstrates the performance across precision, recall, and F-1 score metrics for all classes superior to the baseline and conventional FL. In particular, the precision for particle fault class of the proposed SK-CFL achieves 98.51%, which is 5.55% and 5.65% higher than that of the conventional FL and baseline method, respectively. The recall



(a)

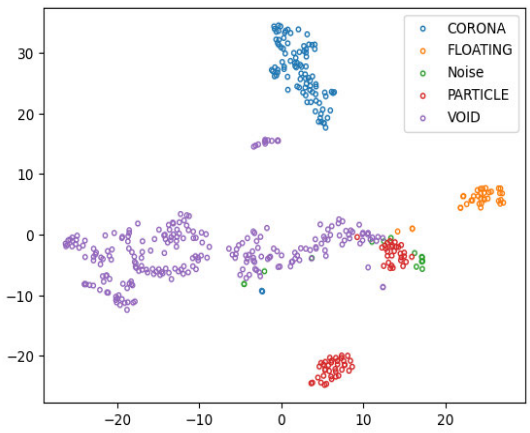


(b)

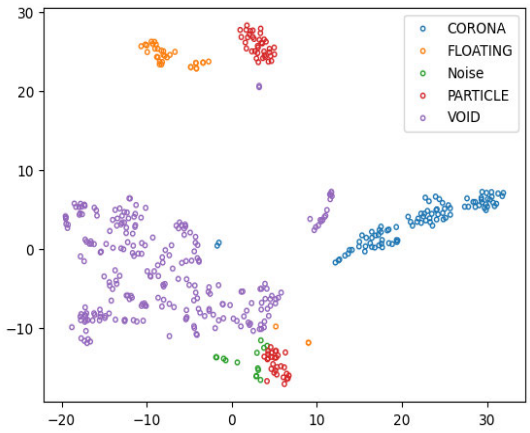


(c)

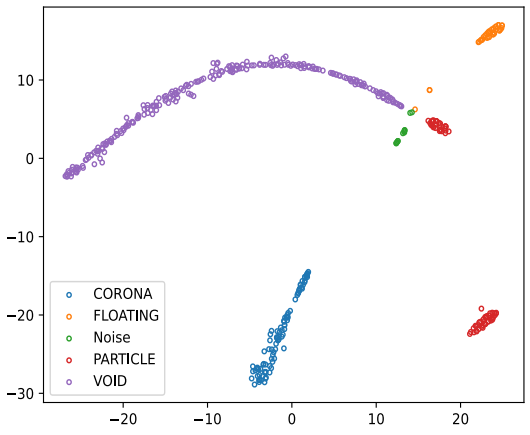
FIGURE 8. Confusion matrix on the untrained IDE 6 (a) baseline, (b) conventional FL, and (c) proposed SK-CFL.



(a)



(b)



(c)

FIGURE 9. t-SNE representations on the untrained IDE 6 for (a) baseline, (b) conventional FL, and (c) proposed SK-CFL.

for noise patterns shows 92.86% of the proposed SK-CFL, compared to 64.29% and 35.71% of the conventional FL and baseline method, respectively. The F-1 score reflects the trade-off of precision and recall, showing that their performance for the proposed model outperforms the conventional FL and baseline method. These results indicate the effectiveness of the proposed approach for various categories, especially in case of imbalanced data.

V. CONCLUSION

In this study, we proposed a SK-CFL method to detect PRPDs in GISs. The proposed SK-CFL utilizes local knowledge

and integrates contrastive learning with federated learning in several IED environments. The proposed SK-CFL was verified when testing a new IED based on PRPDs and on-site noises using multiple local IEDs. The PRPD data from artificial cells included four types of faults: corona, floating, particle, and void. Experimental results have revealed that the proposed SK-CFL not only maintains high classification accuracy within the testing set of the trained local IED but also enhances the fault diagnosis performance in new IEDs. The proposed SK-CFL aggregates locally learned knowledge, thereby resulting in comprehensive generalizations through contrastive learning, which is advantageous when operating in various environments. By training models on local IEDs, the proposed SK-CFL makes it more suitable for real-world industrial scenarios in comparison with the traditional centralized training techniques. In future studies, we intend to conduct PRPD experiments with various gas pressures and obtain more measurements for fault data to verify the proposed method.

REFERENCES

[1] H. D. Ilkhechi and M. H. Samimi, "Applications of the acoustic method in partial discharge measurement: A review," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 28, no. 1, pp. 42–51, Feb. 2021.

[2] X. Li, W. Liu, Y. Xu, and D. Ding, "Discharge characteristics and detectability of metal particles on the spacer surface in gas-insulated switchgears," *IEEE Trans. Power Del.*, vol. 37, no. 1, pp. 187–196, Feb. 2022, doi: 10.1109/TPWRD.2021.3055533.

[3] Y. R. Yadam, R. Sarathi, and K. Arunachalam, "Numerical and experimental investigations on influence of internal defect parameters on partial discharge induced UHF signals in gas insulated switchgear," *IEEE Access*, vol. 10, pp. 110785–110795, 2022.

[4] Z. Wu, B. Lyu, Q. Zhang, L. Liu, and J. Zhao, "Phase-space joint resolved PD characteristics of defects on insulator surface in GIS," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 27, no. 1, pp. 156–163, Feb. 2020.

[5] S. Zheng and S. Wu, "Detection study on propagation characteristics of partial discharge optical signal in GIS," *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, Art. no. 7005912.

[6] L. Zhang, S. Wang, and L. Sun, "Influence of metal particle on impulse flashover characteristics of insulator in GIS," *High Voltage Eng.*, vol. 44, no. 9, pp. 2807–2814, Dec. 2018.

[7] X. Li, M. Wu, X. Hu, P. Jiang, J. Han, K. Zhao, Z. Li, and Q. Zhang, "Discharge characteristics of insulation defects in GIS under AC/LI superimposed voltage," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 29, no. 5, pp. 2017–2025, Oct. 2022.

[8] X. Han, X. Zhang, R. Guo, H. Wang, J. Li, Y. Li, and M. Zhao, "Partial discharge detection in gas-insulated switchgears using sensors integrated with UHF and optical sensing methods," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 29, no. 5, pp. 2026–2033, Oct. 2022.

[9] J. Ardila-Rey, J. Montaña, B. de Castro, R. Schurch, J. C. Ulson, F. Muhammad-Sukki, and N. Bani, "A comparison of inductive sensors in the characterization of partial discharges and electrical noise using the chromatic technique," *Sensors*, vol. 18, no. 4, p. 1021, Mar. 2018.

[10] M. M. Yaacob, M. A. Alsaedi, J. R. Rashed, A. M. Dakhil, and S. F. Atyah, "Review on partial discharge detection techniques related to high voltage power equipment using different sensors," *Photonic Sensors*, vol. 4, no. 4, pp. 325–337, Dec. 2014.

[11] Y. Wang, J. Yan, Z. Yang, Q. Jing, Z. Qi, J. Wang, and Y. Geng, "A domain adaptive deep transfer learning method for gas-insulated switchgear partial discharge diagnosis," *IEEE Trans. Power Del.*, vol. 37, no. 4, pp. 2514–2523, Aug. 2022.

[12] V.-N. Tuyet-Doan, H.-A. Pho, B. Lee, and Y.-H. Kim, "Deep ensemble model for unknown partial discharge diagnosis in gas-insulated switchgears using convolutional neural networks," *IEEE Access*, vol. 9, pp. 80524–80534, 2021.

[13] X. Peng, F. Yang, G. Wang, Y. Wu, L. Li, Z. Li, A. A. Bhatti, C. Zhou, D. M. Hepburn, A. J. Reid, M. D. Judd, and W. H. Siew, "A convolutional neural network-based deep learning methodology for recognition of partial discharge patterns from high-voltage cables," *IEEE Trans. Power Del.*, vol. 34, no. 4, pp. 1460–1469, Aug. 2019.

[14] 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.

[15] V.-N. Tuyet-Doan, T.-D. Do, N.-D. Tran-Thi, Y.-W. Youn, and Y.-H. Kim, "One-shot learning for partial discharge diagnosis using ultra-high-frequency sensor in gas-insulated switchgear," *Sensors*, vol. 20, no. 19, p. 5562, Sep. 2020.

[16] D. Soh, S. B. Krishnan, J. Abraham, L. K. Xian, T. K. Jet, and J. F. Yongyi, "Partial discharge diagnostics: Data cleaning and feature extraction," *Energies*, vol. 15, no. 2, p. 508, Jan. 2022.

[17] B. Ghimire and D. B. Rawat, "Recent advances on federated learning for cybersecurity and cybersecurity for federated learning for Internet of Things," *IEEE Internet Things J.*, vol. 9, no. 11, pp. 8229–8249, Jun. 2022.

[18] H.-S. Lee, "Device selection and resource allocation for layerwise federated learning in wireless networks," *IEEE Syst. J.*, vol. 16, no. 4, pp. 6441–6444, Dec. 2022.

[19] B. Gu, A. Xu, Z. Huo, C. Deng, and H. Huang, "Privacy-preserving asynchronous vertical federated learning algorithms for multiparty collaborative learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6103–6115, Nov. 2022.

[20] J. Zhang, J. Zhou, J. Guo, and X. Sun, "Visual object detection for privacy-preserving federated learning," *IEEE Access*, vol. 11, pp. 33324–33335, 2023.

[21] Y. Ye, S. Li, F. Liu, Y. Tang, and W. Hu, "EdgeFed: Optimized federated learning based on edge computing," *IEEE Access*, vol. 8, pp. 209191–209198, 2020.

[22] N. I. Mowla, N. H. Tran, I. Doh, and K. Chae, "AFRL: Adaptive federated reinforcement learning for intelligent jamming defense in FANET," *J. Commun. Netw.*, vol. 22, no. 3, pp. 244–258, Jun. 2020.

[23] L. Ahmed, K. Ahmad, N. Said, B. Qolomany, J. Qadir, and A. Al-Fuqaha, "Active learning based federated learning for waste and natural disaster image classification," *IEEE Access*, vol. 8, pp. 208518–208531, 2020.

[24] A. Nguyen, T. Do, M. Tran, B. X. Nguyen, C. Duong, T. Phan, E. Tjiputra, and Q. D. Tran, "Deep federated learning for autonomous driving," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Jun. 2022, pp. 1824–1830.

[25] G. Hua, L. Zhu, J. Wu, C. Shen, L. Zhou, and Q. Lin, "Blockchain-based federated learning for intelligent control in heavy haul railway," *IEEE Access*, vol. 8, pp. 176830–176839, 2020.

[26] 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.*, 2020, pp. 10709–10719.

[27] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum contrast for unsupervised visual representation learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9726–9735.

[28] H. Lee, J. Seol, and S.-G. Lee, "Contrastive learning for unsupervised image-to-image translation," 2021, *arXiv:2105.03117*.

[29] H. Fang and P. Xie, "An end-to-end contrastive self-supervised learning framework for language understanding," *Trans. Assoc. Comput. Linguistics*, vol. 10, pp. 1324–1340, Nov. 2022.

[30] L. Sun, N. Yolwas, and L. Jiang, "A method improves speech recognition with contrastive learning in low-resource languages," *Appl. Sci.*, vol. 13, no. 8, p. 4836, Apr. 2023.

[31] Y. Lin, Q. Wang, X. Yu, Z. Zhang, D. Guo, and J. Zhou, "Towards recognition for radio-echo speech in air traffic control: Dataset and a contrastive learning approach," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 31, pp. 3249–3262, 2023.

[32] S. Bhati, J. Villalba, P. Zelasko, L. Moro-Velazquez, and N. Dehak, "Unsupervised speech segmentation and variable rate representation learning using segmental contrastive predictive coding," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 30, pp. 2002–2014, 2022.

[33] V.-N. Tuyet-Doan, T.-T. Nguyen, M.-T. Nguyen, J.-H. Lee, and Y.-H. Kim, "Self-attention network for partial-discharge diagnosis in gas-insulated switchgear," *Energies*, vol. 13, no. 8, p. 2102, Apr. 2020.

[34] 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.

[35] Ogura Jewel Industry Co., Ltd. *Other Product Groups Utilizing High Precision Machining Technology*. Accessed: Jan. 2024. [Online]. Available: http://www.ogura-indus.co.jp/en/product/section17.php/
[36] 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, pp. 1–8, 2022.
[37] H. McMahan, E. Moore, D. Ramage, S. Hampson, and B. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Statist.*, vol. 54, 2017, pp. 1273–1282.
[38] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 18661–18673.
[39] A. van den Oord, Y. Li, and O. Vinyals, "Representation learning with contrastive predictive coding," 2018, *arXiv:1807.03748*.
[40] V. N. Vapnik, "An overview of statistical learning theory," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 988–999, Sep. 1999.
[41] M. Ragab, Z. Chen, W. Zhang, E. Eldele, M. Wu, C.-K. Kwoh, and X. Li, "Conditional contrastive domain generalization for fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.
[42] T. G. Dietterich, "Approximate statistical tests for comparing supervised classification learning algorithms," *Neural Comput.*, vol. 10, no. 7, pp. 1895–1923, Oct. 1998.
[43] H. Dalianis, "Evaluation metrics and evaluation," in *Clinical Text Mining*. Cham, Switzerland: Springer, 2018, pp. 45–53.
[44] L. van der Maaten and G. Hinton, "Visualizing high-dimensional data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

VO-NGUYEN TUYET-DOAN received the B.S. degree in statistics from Ton Duc Thang University, Ho Chi Minh City, Vietnam, in 2018. She is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, Myongji University, South Korea. Her research interests include statistics, data analysis, pattern recognition, computer vision, image processing, and artificial intelligence techniques.

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 with Korea Electrotechnology Research Institute (KERI). His research interests include condition monitoring, signal processing, and machine learning.

HYUN-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 of Taekwang E&C, Seoul, South Korea. From 2006 to 2010, he was the Director of TEC&CO, Seoul. Since 2017, he has been the Director of Genad System, Jeollanam-do, 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 the Korea Electrotechnology Research Institute (KERI), Gyeonggi-do, 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.