IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 12, 2021, accepted May 21, 2021, date of publication May 31, 2021, date of current version June 9, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3084950

Deep Ensemble Model for Unknown Partial Discharge Diagnosis in Gas-Insulated Switchgears Using Convolutional Neural Networks

VO-NGUYEN TUYET-DOAN[1], HA-ANH PHO[1], BYEONGHO LEE[2], (Associate Member, IEEE), AND YONG-HWA KIM[3], (Member, IEEE)

[1]Department of Electronic Engineering, Myongji University, Yongin 17058, South Korea
[2]Department of Electrical and Computer Engineering, Institute of New Media and Communications (INMC), Seoul National University (SNU), Seoul 08826, South Korea
[3]Department of Data Science, Korea National University of Transportation, Uiwang 16106, South 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 and Energy (MOTIE) of the Republic of Korea under Grant 20179310100050 and Grant 20206910100020.

ABSTRACT Deep neural networks (DNNs) are widely used for fault classification using partial discharges (PDs) to evaluate various electrical apparatuses and achieve high classification accuracy pertaining to trained PD faults. However, there is a risk of false alarm in the case of untrained PD faults because it is difficult for DNNs to predict data that were not included in the training process. In this paper, we research classification problems of unknown classes using PDs in gas-insulated switchgears (GISs) and propose a deep ensemble model to obtain the confidence of output probability and determine thresholds to detect unknown fault classes. The proposed model was verified by real-world phase-resolved PD (PRPD) experiments using online ultra-high frequency (UHF) PD measurement systems. The experimental results show that the proposed model achieves better unknown detection performance for the untrained PD faults and retains the classification performance for the trained PD faults.

INDEX TERMS Fault diagnosis, convolutional neural network (CNN), ensemble model, partial discharges (PDs), gas-insulated switchgear (GIS).

I. INTRODUCTION

With the rapid development of power systems, a large number of devices would be needed to regulate high voltages and currents in order to function properly. High current in the flow may cause excessive heating and thus increase the risk of fire or damage to the equipment. To prevent this excessive current quickly and timely from causing unpredictable damage, gas-insulated switchgears (GIS) are used, which are electromechanical devices consisting of high-voltage components, such as circuit breakers and disconnectors. Current is blocked in a circuit automatically when a potential spike is detected in power grids [1]. However, GISs have internal defects during manufacture, transportation, and assembly [2]. In GISs, partial discharges (PDs) are not only the precursors of potential internal defects but also lead to gradual deterioration of insulation and ultimately break down or lead to surface discharge [3], [4]. Therefore, PD detection and diagnosis for GISs are required to ensure the safety and the reliability of power grid assets [5].

PDs can be considered a potential hazard for insulation, the deterioration of which commonly accompanies a series of physical phenomena, such as electromagnetic (EM) radiation, mechanical vibration, and acoustic waves [6]. There are several methods to detect PDs, such as ultra-high frequency (UHF), current induction, and acoustic emission [7]–[9]. The UHF method measures electromagnetic signals within the frequency range of 300—3000 MHz and has been widely used owing to the advantages of high sensitivity, immunity to external interference, ability to locate PD sources, and recognize fault types [1]. Therefore, this study utilizes the UHF method for the PD measurement system [5].

To analyze PD signals, time-resolved PD (TRPD) and phase-resolved PD (PRPD) were studied. Using TRPD signals, the shape of the discharge pulse through time-domain, frequency-domain, and both time- and frequency-domain

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

features were analyzed to determine preliminary estimates [10]–[15]. The time-domain analysis requires few computing resources and simple algorithmic interface with frequency-domain and time-and-frequency domain techniques [13]. Frequency ranges are required to detect TRPD in the frequency-domain analysis [14]. In the time- and frequency-domain analysis, a combination of time and frequency domain can extract effective information to improve pulse separation [15]. Based on the PRPD measurements, the phase-amplitude-number (ϕ-q-n) pattern of the PRPD was analyzed, where ϕ is the phase angle for the PD pulse, q is the amplitude, and n is the number of pulses. The shapes of the defects were identified by analyzing the number of PD pulses and the maximum amplitude or average amplitude in each phase of the PRPD [16]. Using PRPD features, several studies have applied machine learning techniques to classify PDs, whereas other studies have used artificial neural networks (ANNs), support vector machines (SVMs), and decision trees [17]–[22]. The combination of signal-processing techniques and SVMs, which has the advantage of handling the issues of pulsating or localized shape and a ''dimensionality disaster'' [17]. An improved bagging algorithm was used to recognize PD UHF patterns by carrying out wavelet packet transform with a backpropagation neural network and an SVM [18]. However, the major limitation of SVMs is the sensitivity attached to choose the appropriate parameters of the kernel function [19]. The fuzzy theory has a simple architecture, which is a nonlinear mapping from the input space to the output space; however, its classifying ability is insufficient as it is not likely to take advantage of previous diagnosis results [20]. Although ANN has been widely used for PD diagnosis for decades because of superior learning capabilities in input–output relationships, it is still hindered owing to complex architectures with sophisticated activation functions and a larger number of layers and neurons [21], [22]. This complexity can reach a bottleneck in development, which can increase the training time and result in the removal of gradient and overfitting.

The state-of-the-art deep neural networks (DNNs), which can automatically extract high-level features, have achieved promising results in multiple pattern recognition-based tasks such as computer vision, natural language processing, speech recognition, and text classification [23]–[25]. DNNs have been proposed to improve fault classification accuracy using PRPDs [26]–[29]. A deep convolutional neural network (CNN) automatically extracts features using UHF signals for power transformers [26]. Recurrent neural networks (RNNs) with long short-term memory (LSTM) classify PRPD faults in GIS using temporal dependencies from PRPDs [27]. The self-attention model exhibited better performance than the LSTM RNN model by capturing the relevance among the phases of the PDs based on multi-head self-attention and improved computation using parallelism [28]. The stacked sparse autoencoder extracts meaningful features from the PD data for PD pattern classification [29]. However, previous studies have analyzed classification accuracy in a pre-specified class scope [26]–[29], while there is an uncertainty problem of unknown class data.

A common major shortcoming in the existing studies is that the DNN-based classifiers only focus on improving the classification capability for pre-specified targets while ignoring or being incapable to deal with unknown targets. Analyzing the classification performance of a DNN-based classifier is as necessary for untrained data as that for trained data. In addition, acquiring all the types of faults for trained data is a challenging task for PRPD fault diagnosis in GIS as the fault types are very diverse on-site [30]. Consequently, a fault that is not a part of the training data may occur in on-site measurements. Therefore, a fault diagnosis system that detects learned classes and determines unknown classes is worth researching. In general, most classifiers based on neural networks obtain high classification accuracy for trained fault classes but tend to show overconfident prediction for unwanted fault classes suffering from the overfitting problem [31].

The unknown target classifications have been proposed for radar systems [32], [33]. Statistical analysis is performed for unknown training class samples [32]. Prior to classification, a binary hypothesis test is used to detect unknown targets [33]. However, target classification based on deep neural networks has the advantages of extracting features from raw data compared with statistical analysis.

In order to overcome the problem of overconfidence predictions of untrained on-site PRPD data in GISs, we propose a deep ensemble model to identify uncertain faults as an unknown class. The proposed model considers uncertainty estimates, reduces overconfidence predictions, and is efficient at unknown classification. The proposed model comprises multiple CNNs and an ensemble output. The CNN can extract inherent characteristics without any human superintendence by exploiting spatial or temporal correlations in the data owing to the use of multiple feature extraction stages. In comparison with its predecessors, the complexity and difficulty during the training of the CNNs are drastically diminished due to sharing parameters and local connections, which prevent the risk of overfitting [34]. The output of a single CNN is considered as a confidence score to identify unknown classes. Confidence scores obtained from multiple networks are combined for a final prediction.

The major contributions of this study are as follows:

- We propose a deep ensemble model to classify the targets of untrained PRPD faults as unknown fault classes. In the proposed method, we develop a confidence threshold for unknown fault classification based on deep CNNs, which were not investigated in previous PRPD-related studies.
- To the best of our knowledge, this study is the first one which demonstrates that a deep ensemble model can enhance the performance of unknown PRPD diagnosis while maintaining the classification accuracy of the trained PRPDs in the context of GIS. This is because the proposed model can help reduce confidence predictions for unknown test examples during classification [31].

- The proposed model was tested and verified by on-site PRPD experiments in real-world environments using a UHF sensor. The performance was verified by reflecting the on-site PRPD data of seven types of faults that could occur in the GIS, i.e., crack, floating, free particle, void, protrusion on conductors (POCs), protrusion on enclosures (POEs), and particle on spacers (POSs).

The remainder of this paper is organized as follows. In Section II, we introduce the experimental setup and elaborate the results for PRPDs in GISs. The proposed deep ensemble model architecture is presented in Section III. Performance evaluations of the proposed method are presented in Section IV. Finally, Section V concludes the paper.

II. PRELIMINARIES

In this section, we present the experimental setup and experimental results for PRPDs using online UHF PD monitoring systems for GISs (GM-3000) [35]. Here, we use on-site PRPD data of 2003–2015 obtained using the UHF PD monitoring system in GIS [30].

Fig. 1 shows a block diagram of the PD monitoring system, which is composed of a GIS, an internal UHF sensor, and a data acquisition system (DAS). The DAS uses a peak detector to capture the maximum values of the UHF sensor, and N samples in each power cycle were used to measure PRPD. The internal UHF sensor is used to measure PD signals; it has an operating frequency range of 0.5–1.5 GHz and a sensitivity level gain of -14.5 dBm at 5 pC. The sensitivity verification was performed by CIGRE TF 15/33.03.05 [30].

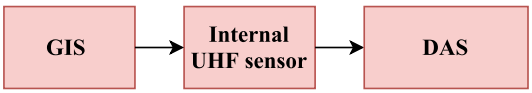

FIGURE 1. Block diagram of the PD monitoring system.

On-site PRPD measurements were conducted using a UHF sensor in a GIS system, where seven types of insulation defects are considered, i.e.., cracks, floating, free particles, voids, POCs, POEs, and POSs. These seven types of insulation defects are described in [30], where void PDs had occurred in 170 kV GIS, floating PDs had occurred in 362 kV GIS, and other PDs had occurred in 275 kV GIS. The measured signal is defined as

$$X = \left[x_1^T, x_2^T, \ldots, x_M^T\right]^T, \tag{1}$$

where $x_m = [x_{m1}, x_{m1}, \ldots, x_{mN}]$, $m = 1, \ldots, M$, N is the number of phase angles, and M is the number of power cycles.

Fig. 2 shows PRPDs for seven fault types, for a PRPD signal of $M = 1000$ power cycles and $N = 256$ phase angles. Fig. 2a shows that the discharge pulse of the crack is observed at both halves of the cycle with an amplitude of -50 dBm, which has a distribution and an amplitude similar to that of free particle PDs. However, the frequency of pulses in free particle PDs is lesser. Fig. 2b shows that the PD pulses for floating are sparsely distributed at both a positive and negative

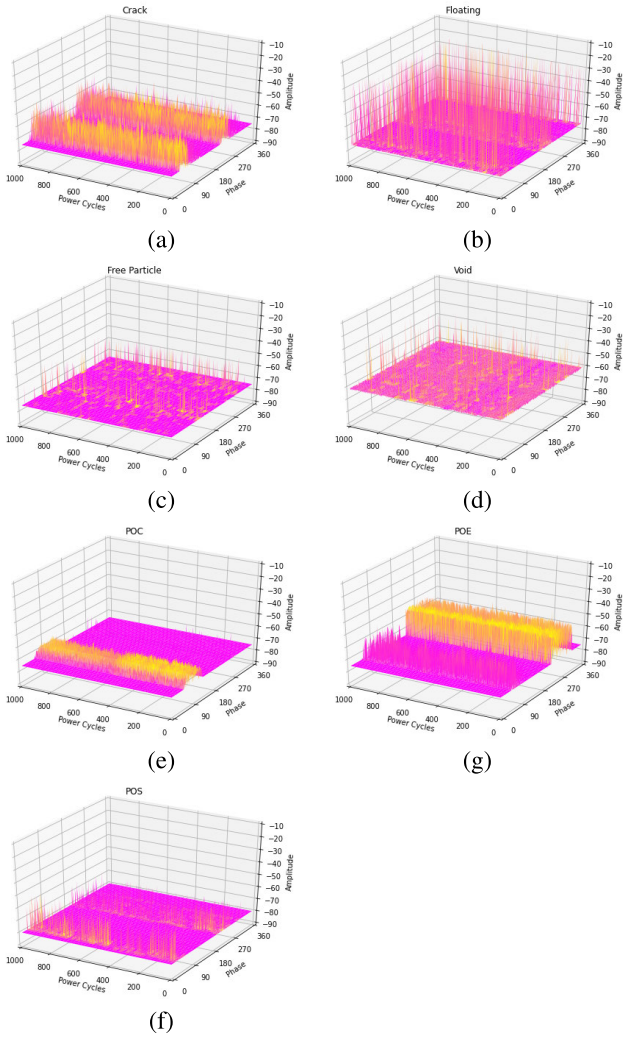

FIGURE 2. PRPDs for seven fault types in GIS: (a) crack, (b) floating, (c) free particle, (d) void, (e) POC, (g) POE, and (f) POS.

half cycle, with a maximum amplitude of -30 dBm. For the void PDs, the discharge pulse across all bands at varying intensity ranges from -65 to -55 dBm, as shown in Fig. 2d. In the POC, PD pulses occurred mostly in the region around 90–180 in the negative band. The discharge pulses of POE are mainly distributed from 270 to 360. PDs for POS are mostly observed around regions 0–90 and 180–300, with a sparse occurrence in a range of -75 to -50 dBm.

Fig. 3 depicts the t-distributed stochastic neighbor embedding (t-SNE) representation of vectors for the inputs of seven PD fault types. The t-SNE deployed high-dimensional vectors into 2D spaces while retaining local similarities or pairwise distances and providing a visual indication of the data arrangement in a high-dimensional space [36]. As can be seen in Fig. 3, the distributions of raw PD signals are very close to each other; therefore, it is difficult to distinguish discharge faults precisely based on the raw PD signals.

Fig. 4 depicts the amplitude levels of PD for on-site PRPD measurements utilizing statistical parameters such

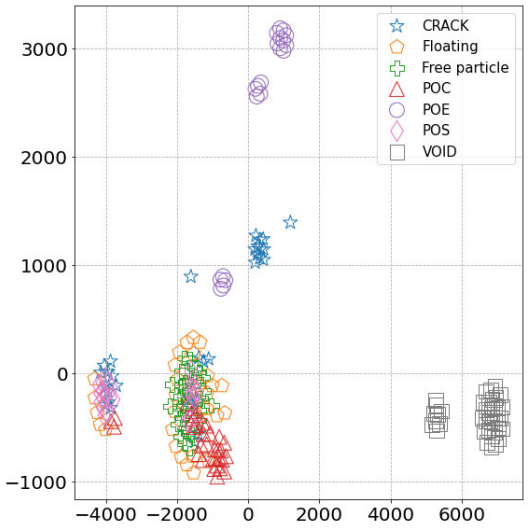

FIGURE 3. Visualization of feature representations for the raw input via t-distributed stochastic neighbor embedding (t-SNE).

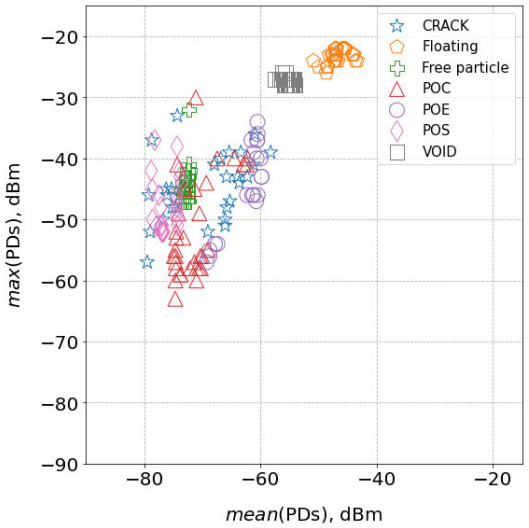

FIGURE 4. Amplitude levels of PD for on-site PRPD measurements utilizing statistical parameters.

as $mean(PDs)$ and $max(PDs)$, in which $mean(PDs)$ and $max(PDs)$ are calculated as $mean(PDs) = mean(X) = \frac{\sum_{m=1}^{M}\sum_{n=1}^{N} x_{mn}}{MN}$ and $max(PDs) = max(X) = \max\{x_{11}, x_{12}, \ldots, x_{MN}\}$. Here, we have considered crack, floating, free particle, and void as known PDs; POC, POE, and POS are considered as unknown PDs. The mean(PDs) of known PDs is $[-80, -43]$ dBm, max(PDs) of known PDs is $[-58, -21]$ dBm, mean(PDs) of unknown PDs is $[-80, -60]$ dBm, and max(PDs) of unknown PDs is $[-63, -30]$ dBm. Therefore, it is difficult to classify known and unknown PDs based on the amplitude levels of PDs.

III. PROPOSED SCHEME

In this section, we define the uncertainty estimation problem for fault diagnosis and describe the architecture of the proposed method, intended to detect unknown classes in a GIS. The proposed model employs a deep ensemble structure using multiple CNN models and determines the threshold using on-site PRPDs to estimate unknown classes.

A. PROBLEM FORMULATION

For fault diagnosis, a large amount of training data is required for each fault type. However, it is difficult to obtain all types of real-world fault data. To prevent unexpected insulation failures in GIS systems, it is necessary to detect known and unknown PRPDs at an early stage. In this study, we focus on developing an uncertainty estimation system to classify unknown classes using real-world PRPD fault data.

A CNN automatically extracts features from PRPDs [26]. However, there is a shortcoming in extracting features from a single CNN model. In particular, the problem arises due to misclassification of unknown classes (classes that do not appear in the training data) as known classes with high confidence. Therefore, when CNNs are tested on a dataset with a mixture of known and unknown classes, it can yield inaccurate predictions. In this study, we classify untrained PRPD fault data into an unknown class based on a deep ensemble model using uncertainty estimation.

Let $f(\mathrm{X}) : \mathbb{R}^{M\times N} \rightarrow \mathbb{R}^{K}_{(0,1)}$ denote a function mapping an input matrix X to a vector that provides a probability distribution consisting of K probabilities proportional to K consistently trained classes; M denotes the power cycle, N is the phase angle, and $\mathbb{R}^{K}_{(0,1)} = \{x_i \in (0,1) | i = 1, \ldots, K\}$ is the set in which each element lies in the interval $(0, 1)$ and the elements add up to 1. The highest value is considered the confidence evaluation γ, whereas the lower confidence evaluation indicates a higher chance of the corresponding data point being unknown. We define the confidence threshold Th_c to detect unknown samples with $\gamma < Th_c$.

Our goal is to study the confidence score γ and the corresponding confidence threshold Th_c, to obtain the best detection accuracy of unknown faults on the test data while retaining the classification accuracy for the trained PD data.

B. DEEP ENSEMBLE MODEL USING CNN

To estimate uncertainty, we propose a deep ensemble network, which consists of an input, multiple CNNs, and an output. The overall architecture of the proposed deep ensemble neural network is shown in Fig. 5. The input is a sequential PRPD signal $X = {}^{M\times N}$ in (1).

In the ensemble network, each CNN consists of two convolutional layers, two max-pooling layers, two dropout layers, a flattened layer, two fully connected layers, and an output layer. The convolutional layer has 64 filters with kernel size of 3×3 and adopts the rectified linear unit activation function to improve the speed of the backpropagation and reduce the vanishing gradients [37]. We set the kernel size of max-pooling to 2, to reduce the parameters of the networks and prevent over-fitting. Dropout is used for regularization [38]. Furthermore, the three-dimensional matrix obtained at this stage is converted to a vector through a flattened layer. There are two fully connected layers in each CNN, with the number of nodes

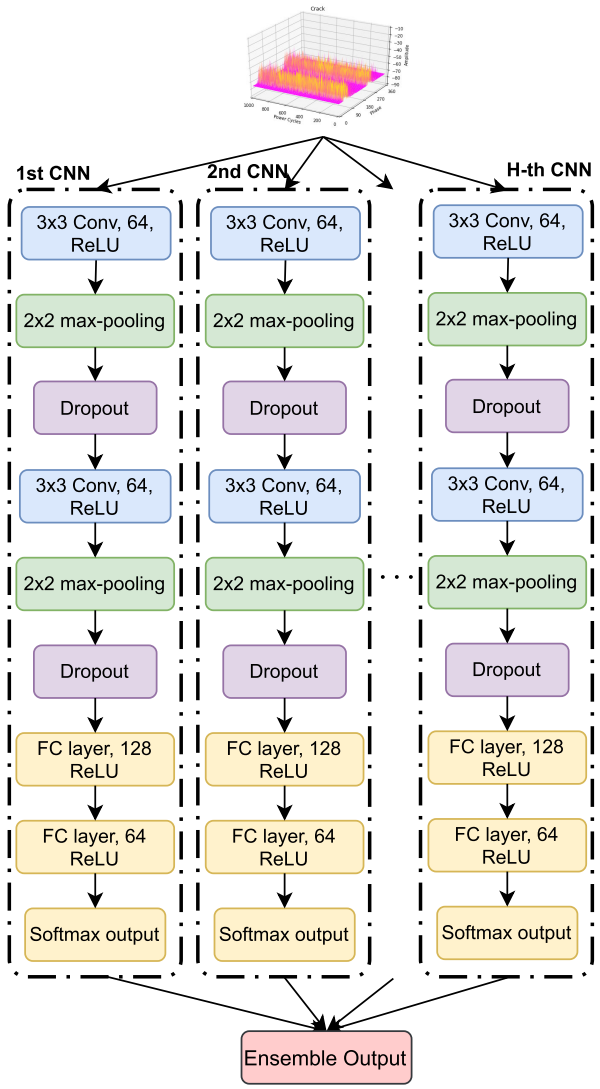

FIGURE 5. Architecture of the proposed deep ensemble model.

of each fully connected layer being 128 and 64, respectively. The output layer represents the probability of each class and is derived through a softmax function [39], which is defined as

$$\sigma_k = \frac{e^{z_k}}{\sum_{j=1}^{K} e^{z_j}}, \tag{2}$$

where $k = 1, \cdots, K$ is the index of the target class, the total number of classes is K, and z_k is the input of the activation function. The softmax output in (2) is a relative probability value in the trained labels. The ensemble output is defined as follows

$$\sigma_{k,ensemble} = \frac{1}{H} \sum_{h=1}^{H} \sigma_{k,h}, \tag{3}$$

where $\sigma_{k,h}$ is the softmax output of the h-th CNN network and H is the number of CNN networks. In (3), the untrained classes become one of the trained classes without proper thresholding, although they have a low similarity score.

To determine the uncertainty of the unknown classes, we set up the confidence value γ, which can help the model identify whether the data are from a known label or not. The confidence value is defined as

$$\gamma = \max\left(\sigma_{k,ensemble}\right), \tag{4}$$

where $k = 1, \cdots, K$. The confidence value γ can determine the input PRPD as an unknown or known fault according to a threshold value Th_c. If $\gamma < Th_c$, the output is an unknown class, and the output is a trained label otherwise. Ensemble size can be determined according to the input data characteristics and the target in classification. In addition, the threshold value Th_c was determined through our experiments, and the most optimal threshold value was used. The details of determining the ensemble size and the threshold, Th_c, are described in Section IV.

C. TRAIN AND TEST FOR NETWORK

The parameters of each CNN were learned through mini-batch training, and the loss function was minimized. Cross-entropy was used as the loss function; it is defined as

$$Loss(b) = -\sum_{k=1}^{K} t_k \log\left(\sigma_k^{(b)}\right), \tag{5}$$

where $\sigma_k^{(b)}$, t_k is denoted as 1 when t_k is a true index and t_k is denoted as 0 otherwise, for the b-th training data in a mini-batch $\mathcal{B}$. The total loss J is calculated as

$$J = \frac{1}{|\mathcal{B}|} \sum_{b \in \mathcal{B}} Loss(b). \tag{6}$$

To train the model, the Adam optimizer was used with an initial learning rate of 0.002 [40]. The overall process of training and test phases of the proposed deep ensemble model is summarized in **Algorithm 1**.

IV. PERFORMANCE EVALUATION

In this section, we present on-site PD measurements for GIS using the online UHF PD measurement system and the performance evaluation of the deep ensemble model for PRPD diagnosis. To detect untrained PD patterns, we consider cracks, floating, free particles, and voids as known classes and POCs, POEs, and POSs as unknown classes. Table 1 shows the number of measurements obtained for the PRPD faults, where the number of PRPDs for known classes is 137 and the number of PRPDs for the unknown class is 78. For the known classes, PRPDs were divided into three parts: training set, validation set, and test set; these were 50%, 20%, and 30% of the original set, respectively. In addition, PRPDs for the unknown class were used as the test set. Therefore, the size of the training set, validation, and test set were 76, 19, and 120, respectively.

In our experiments, we used randomized stratified sampling to ensure that all models were represented equally in the training and test sets. All of these sets were separate. We deployed extensive experiments to acquire optimized hyperparameters for the different parameters used to tune our model. Some hyperparameters are related to the neural

Algorithm 1 Deep Ensemble Algorithm for Classification

Initialize: Labelled training, validation, and test sets X in (1);
<Training phase>:
for $h = 1 : H$ **do**
  Establish h-th CNN network C_h with randomly selected weights and biases;
  Train C_h independently with the training set;
  Loop: **for** all training data
  Step 1: The output of the softmax layer in C_h is calculated as in (2);
  $\sigma_{k,h} = \frac{e^{z_{k,h}}}{\sum_{j=1}^{K} e^{z_{j,h}}}$;
  Step 2: Calculate the error J in (6);
  $Loss(b) = -\sum_{k=1}^{K} t_k \log\left(\sigma_{k,h}^{(b)}\right)$,
  $J = \frac{1}{|\mathcal{B}|} \sum_{b \in \mathcal{B}} Loss(b)$;
  Step 3: Update C_h parameters using back propagation;
  End for
  Find the optimal parameters of C_h;
end for
<Test phase>:
Set a confidence threshold Th_c;
for $h = 1 : H$ **do**
  Based on the C_h model, parameters are optimized, and the output of the softmax layer is calculated.
end for
Obtain the output ensemble $\sigma_{k,ensemble}$ in (3);
Evaluate the confidence γ in (4);
Predict the labels of the test set by taking the maximum probability of the output ensemble
label $= \arg\max_k(\sigma_{k,ensemble})$.
if $\gamma > Th_c$ **then**
  Output: label
else
  Output: unknown class decision
end if

TABLE 1. Experimental dataset for PRPDs in GISs.

	Known class				Unknown class		
Fault type	Crack	Floating	Free Particle	Void	POC	POE	POS
Number of experiments	32	30	33	42	33	20	25

network structure, such as the number of layers, kernel size, and number of kernels. The others are related to the training algorithm, such as learning rate, batch size, and the number of epochs. Table 2 depicts the optimized procession of hyperparameters within the minimum and maximum boundary ranges and indicates whether the parameter value was an integer or a real number. During the executed processes, the 2-layer CNN with a 3×3-size kernel and 64 filters in each layer combine 2-layer FNN with 128 and 64 notes based on the learning rate, drop factor, and batch size, as well as epochs of 0.002, 0.5, 32, and 100, obtaining the highest overall classification accuracy for the trained PD data.

TABLE 2. Hyperparameter optimization via minimum and maximum bound.

Hyperparameter	Minimum	Maximum	Type
Number of layers	1	8	Integer
Kernel size	3×3	8×8	Integer
Number of kernels	16	256	Integer
Learning rate	0.0001	0.01	Real
Drop factor	0.3	0.5	Real
Number of epochs	50	500	Integer

The experiments were programmed in TensorFlow [41], [42] and Keras [43] libraries. TensorFlow is an open-source software library developed by Google for numerical computation using data flow graphs. Computing a model can be easily deployed with the support of CPUs or GPUs. Keras is a high-level API for neural network and is written in Python; it can run on top of TensorFlow. It is the most widely used deep learning framework and focuses on fast analysis.

Fig. 6 shows the training and validation accuracy of a single CNN model. These experimental results were based on known training samples. The single CNN algorithm achieved 100% accuracy for the training set in 20 iterations. Furthermore, the validation dataset could be verified with high accuracy, with only a small difference in performance. Thus, the combination of more CNN models maintains classification accuracy for known classes.

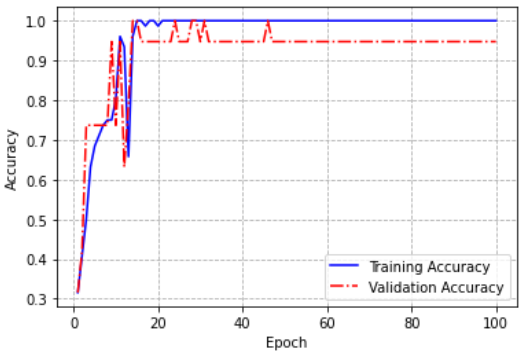

FIGURE 6. Training and validation accuracy of a single CNN model for each epoch.

Fig. 7 shows the normalized confusion matrix of a single CNN model without a confidence threshold for known and unknown classes. Crack, floating, free particle, and void PDs were considered as the known classes. In contrast, POCs, POEs, and POSs did not appear in the training stage, and they were used in the test stage to evaluate the classification performance for unknown classes. As shown in Fig. 7, the single CNN model has 100% accuracy for all the trained faults. However, unknown classes are determined as one of trained classes. POC is determined as a free particle PD. Additionally, 75% of POEs were classified as a crack and 25% as free particles. Moreover, 28% of POSs were determined as a

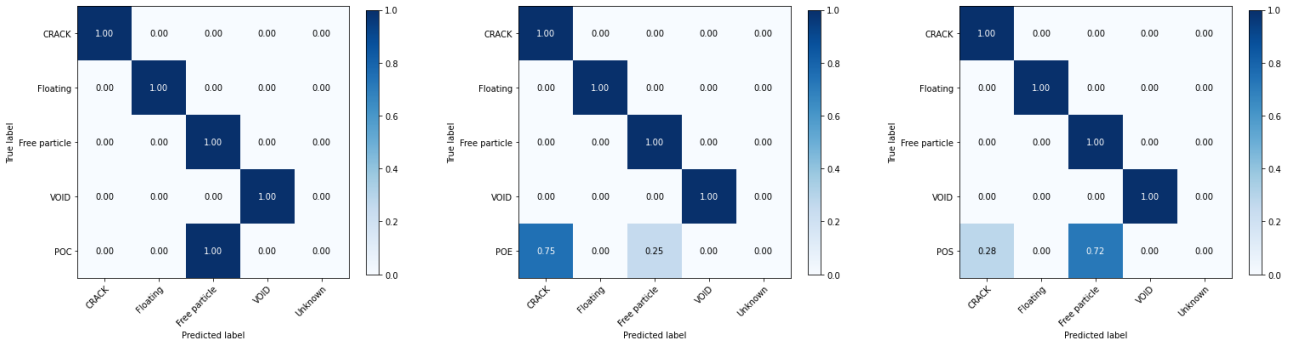

FIGURE 7. Confusion matrix of a single CNN model for known and unknown classes.

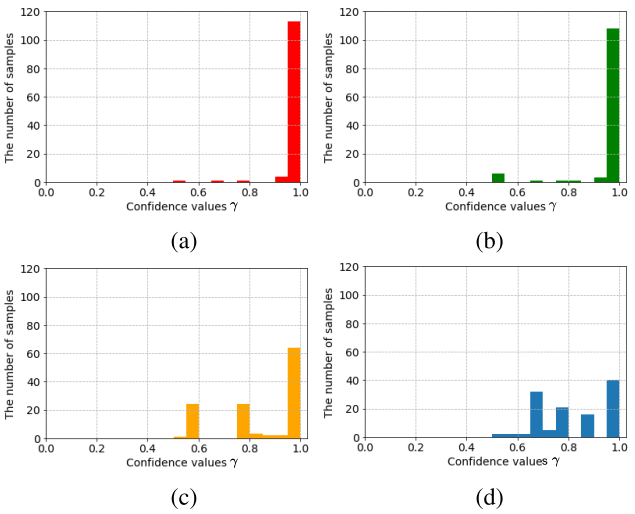

FIGURE 8. Histogram of the output confidence in the test set, including known and unknown class data: (a) single CNN model and deep ensemble model with size, (b) H = 2, (c) H = 5, and (d) H = 10.

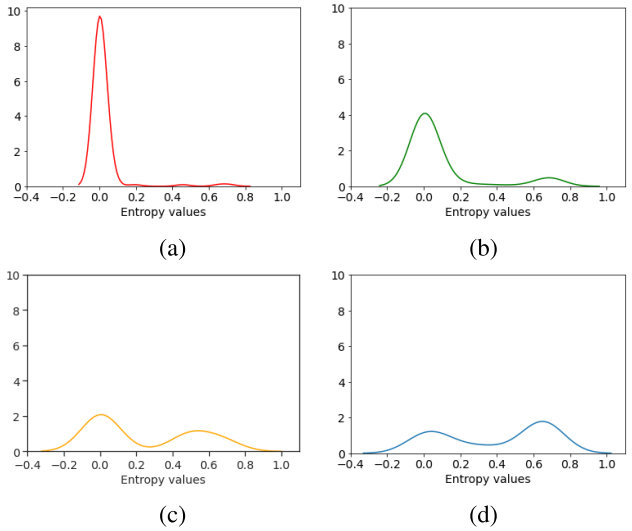

FIGURE 9. Distribution based on the histogram of the predictive entropy in the test set including known class and unknown class data: (a) a single CNN model, deep ensemble model with size (b) H = 2, (c) H = 5, and (d) H = 10.

crack and 72% as a free particle fault. Therefore, it is difficult to determine a fault type by using a single CNN model. The inaccurate prediction and lack of thresholding for the output of the single CNN model can be identified as the root of the problem.

Fig. 8 shows the histogram for the probability of the confidence value γ for the test set, including both known and unknown classes. In Fig. 8a, most samples have high confidence probability (at 100%). This illustrates that the single CNN model is overconfident for both known and unknown class data. As shown in Fig. 8b, c, and d, the deep ensemble model with size $H = 2$, $H = 5$, and $H = 10$, gradually controlled the confidence of the unknown class data when the size increased. As clearly shown in Fig. 8d, there is a distinction between the two regions above and below 90% of confidence.

By averaging the outputs of the independent CNN networks in the ensemble model, the confidence value of unknown classes becomes low. In addition, the confidence values of trained classes remain constant. It can be interpreted that the confidence estimate of the deep ensemble model is more calibrated for the unknown classes.

The quality of uncertainty estimates of CNN and our proposed methods of known and unknown class data are evaluated based on the entropy of the predictive distribution [31]. The entropy value is calculated as [44]

$$Entropy = -\sum_{k=1}^{K} \sigma_{k,ensemble} \log\left(\sigma_{k,ensemble}\right). \quad (7)$$

Fig. 9 shows the estimated kernel distribution based on the histogram of the entropy for a single CNN and the deep ensemble models in the test set. For a single CNN model, the distribution of the entropy value peak is at $Entropy = 0$ with small variance. Meanwhile, as H increases, these distributions significantly change in accordance with prediction confidence and with an increase in entropy values. As shown in Fig. 9d, the density at $Entropy = 0.65$ is greater than that at $Entropy = 0$. This result is suitable for the distribution of unknown classes as the number of unknown classes is larger than that of the known classes in the test set.

TABLE 3. Performance evaluation with different thresholds of related methods.

	Ensemble model with size	$Thc = 0$	$Thc = 0.75$	$Thc = 0.85$	$Thc = 0.95$
Accuracy for **trained classes**	$H = 1$	100%	100%	97.76%	95.24%
	$H = 5$	100%	100%	95.24%	95.24%
	$H = 10$	100%	100%	95.24%	**95.24%**
Accuracy for **untrained classes**	$H = 1$	0%	0%	0%	2.56%
	$H = 5$	0%	29.49%	42.30%	60%
	$H = 10$	0%	75.64%	87.18%	**100%**

TABLE 4. An analysis of the sensitivity and precision of related methods.

	Ensemble model with size	Known class				Unknown class		
		Crack	Floating	Free Particle	Void	POC	POE	POS
Sensitivity analysis	$H = 1$	80%	100%	100%	100%	0%	0%	8%
	$H = 5$	80%	100%	100%	100%	63.64%	30%	80%
	$H = 10$	**80%**	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**
Precision analysis	$H = 1$	28.57%	100%	15.15%	100%	0%	0%	50%
	$H = 5$	36.36%	100%	37.04%	100%	91.30%	75%	66.67%
	$H = 10$	**100%**	**100%**	**100%**	**100%**	**94.29%**	**90.91%**	**92.59%**

Depending on the ensemble size, the threshold for confidence values is determined differently based on the results of the confidence evaluation predictions. The classification accuracy for known and unknown classes involves a trade-off relationship. As shown in Fig. 8d, confidences have a clear difference around 0.95, where lower values can be predicted as an unknown class and higher values are predicted as known classes. Therefore, we set $Th_c = 0.95$ for the deep ensemble model with $H = 10$.

Table 3 illustrates the classification accuracy of trained and untrained class data according to the confidence threshold Thc of various sizes of ensemble model. The ensemble model with size $H = 1$ (the conventional CNN model) shows the degree of trained data when increasing the threshold value is relatively high, with accuracy rates of 100% for $Thc = 0$ and 0.75; 97.76% and 95.24% for $Thc = 0.85$ and 0.95, respectively. However, when the classification is performed for untrained data, the thresholds do not improve the categorization accuracy. The performance of the trained class data of ensemble models with size $H = 5$ and 10 still remains constant based on the different confidence threshold Thc. Moreover, the accuracy ratios of the deep ensemble model with $H = 5$ and 10 for untrained fault diagnosis increase with an increase in the threshold values. In this study, the deep ensemble model with size $H = 10$ is selected. When the accuracy rate is secured at 0.95, the decision performance for the untrained target achieves 100%, which is an improvement of 97.44% compared with that of the conventional CNN model.

The sensitivity and precision of the proposed methods are listed in Table 4, defined as

$$sensitivity = \frac{TP}{TP + FN}, \tag{8}$$

$$precision = \frac{TP}{TP + FP}, \tag{9}$$

where TP is True Positive, FN is False Negative, and FP is False Positive [45]. Sensitivity measures the percentage of samples that are correctly identified out of all positive predictions that could have been made, which indicates the coverage of predictions of a class. As can be seen in Table 4, the proposed ensemble model with size $H = 10$ gained 100% efficiency for unknown fault classes, while single CNN did not classify unknown PD classes in 78 actual samples of unknown data. Precision evaluates the percentage of correct results in all positive predictions of a class, which shows the accuracy of each prediction our model made to that class. The precision of ensemble with size $H = 10$ in each of the known classes is 100%. Here, there is not a class that is predicted to confuse with them even unknown classes. Moreover, the precision of the proposed ensemble method for unknown classes is above 90% because there is a small sample of known classes, which has a confidence value below $Thc = 0.95$; therefore, it is predicted to be an unknown class.

Fig. 10 shows the confusion matrix of the proposed deep ensemble model with $H = 1$, $H = 5$, and $H = 10$ for the test set. For better comparison, the thresholds are determined with 0.95 for all cases, and the red rectangle highlights the test set performance. Fig. 10a illustrates the result of the single CNN model. In this case, the crack defect is mistakenly recognized as an unknown fault with 20% probability, and the unknown classes POC and POS are misclassified as cracks and free particles. As shown in Fig. 10b, the performance of the ensemble with size $H = 5$ remains up to par for trained classes, but the accuracy for unknown (POC, POE, and POS) classes is 64%, 30%, and 80%, respectively. As shown in Fig. 10c, the performance of the proposed ensemble with size $H = 10$ clearly achieves 100% classification accuracy for unknown classes. Therefore, the proposed method can solve the problem of determining unknown faults as an unknown class. As it is difficult to obtain all types of fault data in the real world in advance, the proposed method can improve fault classification performance by acquiring unknown fault PRPDs in GISs.

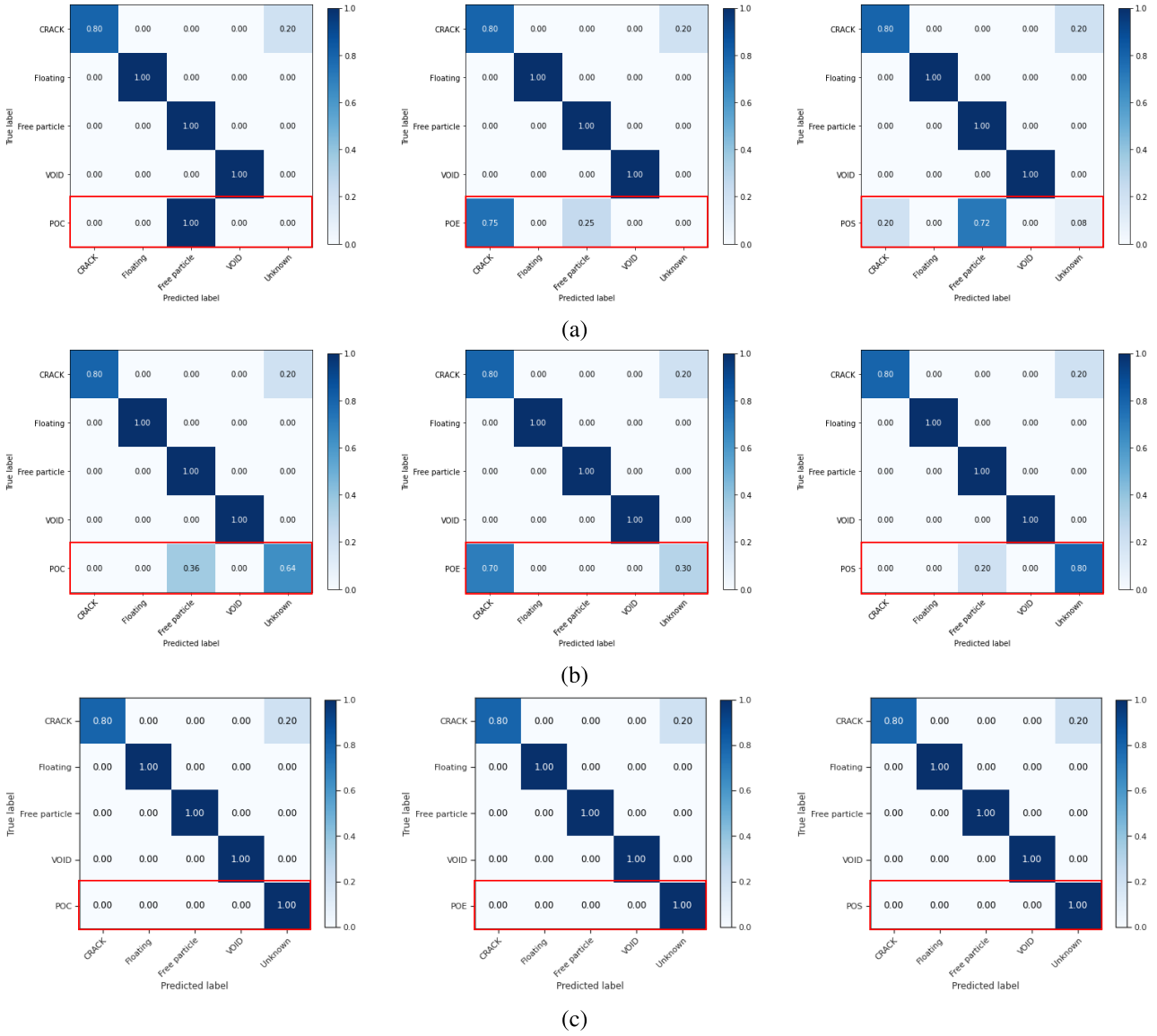

FIGURE 10. Confusion matrix of deep ensemble model with threshold = 95% for different sizes: (a) H = 1, (b) H = 5, and (c) H = 10. The first column is the unknown class performance for POC, the second one is for POE, and the third one for POS.

V. CONCLUSION

Deep neural networks have gained attention as a state-of-the-art technology for fault diagnosis in power equipment. However, there are overconfidence problems associated with classification of untrained data. To prevent incorrect predictions on untrained data, we propose a deep ensemble model based on uncertainty estimation for fault diagnosis in GISs. The proposed method considers untrained data classes as an unknown class. Using real-world on-site measurements, experimental results show that the proposed method detects the unknown PRPD faults with 100% accuracy while retaining the classification performance for trained PRPD faults. Moreover, other unknown faults that can occur on-site can be classified using the proposed deep ensemble model. This is because the determination of the threshold for the confidence values does not depend on the type of unknown faults; rather it depends on the ensemble size of the proposed deep ensemble model. The proposed method can be used to in the maintenance of GISs when detecting PRPDs. In addition, if unknown PRPDs are detected by the proposed method, we can investigate new fault types and their corresponding PRPD data for a stable power grid operation and grid asset management system.

In future studies, we intend to obtain various on-site faults to evaluate the proposed method, analyze different neural networks considering different hyperparameters for each CNN for the deep ensemble model, and conduct further verifications of the proposed method for life degradation based on the effect of each level of PD and for the severity levels of PD.

REFERENCES

[1] M. M. Rao and M. Kumar, "Ultra-high frequency (UHF) based partial discharge measurement in gas insulated switchgear (GIS)," in *Proc. Int. Conf. High Voltage Eng. Technol. (ICHVET)*, Hyderabad, India, Feb. 2019, pp. 1–5.

[2] W. Boeck, "Insulation Co-ordination of GIS, return of experience, on site tests and diagnostic techniques," *Electra*, vol. 176, no. 2, pp. 67–95, 1998.

[3] F. Zeng, J. Tang, X. Zhang, S. Zhou, C. Pan, and R. A. Sánchez, "Typical internal defects of gas-insulated switchgear and partial discharge characteristics," in *Simulation and Modelling of Electrical Insulation Weaknesses in Electrical Equipment*, R. A. Sánchez, Ed. Rijeka, Croatia: InTech, 2018.

[4] Y. Li, T. Jin, T. Wang, G. Yang, and Y. Li, "Research on internal typical partial discharge characteristics for gas insulation switchgear," in *Proc. Chin. Control Decis. Conf. (CCDC)*, Shenyang, China, Jun. 2018, pp. 5166–5170.

[5] X. Li, H. Tang, S. Mu, K. Song, R. Liu, G. Xu, and Q. Li, "Partial discharge monitoring system for PD characteristics of typical defects in GIS using UHF method," in *Proc. Int. Conf. High Voltage Eng. Appl.*, Shanghai, China, Sep. 2012, pp. 625–628.

[6] I. A. Metwally, "Status review on partial discharge measurement techniques in gas-insulated switchgear/lines," *Electr. Power Syst. Res.*, vol. 69, no. 1, pp. 25–36, Apr. 2004.

[7] S. Tenbohlen, D. Denissov, S. Hoek, and S. M. Markalous, "Partial discharge measurement in the ultra high frequency (UHF) range," *IEEE Trans. Dielectrics Electr. Insul.*, vol. 15, no. 6, pp. 1544–1552, Dec. 2008.

[8] *PD IEC/TS 62478:2016–High-Voltage Test Techniques: Measurement of Partial Discharge by Electromagnetic and Acoustic Methods*, International Electrotechnical Commission, Geneva, Switzerland, 2016.

[9] L. E. Lundyaard, "Partial discharge—Part XIV: Acoustic partial discharge detection-practice application," *IEEE Elect. Insul. Mag.*, vol. 8, no. 5, pp. 34–43, Sep./Oct. 1992.

[10] R. P. Nair and S. B. Vishwanath, "Analysis of partial discharge sources in stator insulation system using variable excitation frequency," *IET Sci., Meas. Technol.*, vol. 13, no. 6, pp. 922–930, Aug. 2019.

[11] X. Chen, Y. Qian, G. Sheng, and X. Jiang, "A time-domain characterization method for UHF partial discharge sensors," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 24, no. 1, pp. 110–119, Feb. 2017.

[12] H. Jahangir, A. Akbari, P. Werle, M. Akbari, and J. Szczechowski, "UHF characteristics of different types of PD sources in power transformers," in *Proc. Iranian Conf. Elect. Eng. (ICEE)*, May 2017, pp. 1242–1247.

[13] C. Chang, C. S. Chang, J. Jin, T. Hoshino, M. Hanai, and N. Kobayashi, "Source classification of partial discharge for gas insulated substation using waveshape pattern recognition," *IEEE Trans. Dielectrics Electr. Insul.*, vol. 12, no. 2, pp. 374–386, Apr. 2005.

[14] D. Denissov, W. Kohler, S. Tenbohlen, R. Grund, and T. Klein, "Wide and narrow band PD detection in plug-in cable connectors in the UHF range," in *Proc. Int. Conf. Condition Monitor. Diagnosis*, Beijing, China, Apr. 2008, pp. 1056–1059.

[15] A. Contin, A. Cavallini, G. C. Montanari, G. Pasini, and F. Puletti, "Digital detection and fuzzy classification of partial discharge signals," *IEEE Trans. Dielectrics Electr. Insul.*, vol. 9, no. 3, pp. 335–348, Jun. 2002.

[16] B. Karthikeyan, S. Gopal, and S. Venkatesh, "Partial discharge pattern classification using composite versions of probabilistic neural network inference engine," *Expert Syst. Appl.*, vol. 34, no. 3, pp. 1938–1947, Apr. 2008.

[17] H. D. O. Mota, L. C. D. D. Rocha, T. C. D. M. Salles, and F. H. Vasconcelos, "Partial discharge signal denoising with spatially adaptive wavelet thresholding and support vector machines," *Electr. Power Syst. Res.*, vol. 81, no. 2, pp. 644–659, Feb. 2011.

[18] T. Jiang, J. Li, Y. Zheng, and C. Sun, "Improved bagging algorithm for pattern recognition in UHF signals of partial discharges," *Energies*, vol. 4, no. 7, pp. 1087–1101, Jul. 2011.

[19] S. Lu, H. Chai, A. Sahoo, and B. T. Phung, "Condition monitoring based on partial discharge diagnostics using machine learning methods: A comprehensive state-of-the-art review," *IEEE Trans. Dielectrics Electr. Insul.*, vol. 27, no. 6, pp. 1861–1888, Dec. 2020.

[20] T. K. Abdel-Galil, R. M. Sharkawy, M. M. A. Salama, and R. Bartnikas, "Partial discharge pattern classification using the fuzzy decision tree approach," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 6, pp. 2258–2263, Dec. 2005.

[21] A. A. Mas'ud, R. Albarracín, J. A. Ardila-Rey, F. Muhammad-Sukki, H. A. Illias, N. A. Bani, and A. B. Munir, "Artificial neural network application for partial discharge recognition: Survey and future directions," *Energies*, vol. 9, no. 8, pp. 574–591, 2016.

[22] E. Gulski and A. Krivda, "Neural networks as a tool for recognition of partial discharges," *IEEE Trans. Electr. Insul.*, vol. 28, no. 6, pp. 984–1001, Dec. 1993.

[23] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[24] A. Hassan and A. Mahmood, "Convolutional recurrent deep learning model for sentence classification," *IEEE Access*, vol. 6, pp. 13949–13957, 2018.

[25] G. Chen, S. Zhang, X. Tao, and X. Zhao, "Speech emotion recognition by combining a unified first-order attention network with data balance," *IEEE Access*, vol. 8, pp. 215851–215862, 2020.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] S.-W. Kim, J.-R. Jung, Y.-M. Kim, G.-S. Kil, and G. Wang, "New diagnosis method of unknown phase-shifted PD signals for gas insulated switchgears," *IEEE Trans. Dielectrics Electr. Insul.*, vol. 25, no. 1, pp. 102–109, Feb. 2018.

[31] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2017, pp. 6402–6413.

[32] G. E. Smith, M. Vespe, K. Woodbridge, and C. J. Baker, "Radar classification evaluation," in *Proc. IEEE Radar Conf.*, Rome, Italy, May 2008, pp. 1–6.

[33] T. Haumtratz, J. Worms, and J. Schiller, "Classification of air targets including a rejection stage for unknown targets," in *Proc. 11th Int. Radar Symp.*, Vilnius, Lithuania, Jun. 2010, pp. 1–4.

[34] S.-H. Bae and K.-J. Yoon, "Confidence-based data association and discriminative deep appearance learning for robust online multi-object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 595–610, Mar. 2018.

[35] Hyosung Corporation. *Hyosung Intelligent Condition Monitoring System Catalogue*. Accessed: Apr. 16, 2021. [Online]. Available: http://www.hyosungheavyindustries.com/en/business/monitoringsystem.do

[36] 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.

[37] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

[38] L. I. Kuncheva, C. J. Whitaker, C. A. Shipp, and R. P. W. Duin, "Limits on the majority vote accuracy in classifier fusion," *Pattern Anal. Appl.*, vol. 6, no. 1, pp. 22–31, Apr. 2003.

[39] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[40] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980

[41] M. Abadi *et al.*, "Tensorflow: A system for large-scale machine learning," *12th USENIX Symp. Operating Syst. Design Implement. (OSDI)*, vol. 16, 2016, pp. 265–283.

[42] (2016). *Tensorflow: An End-to-End Open Source Machine Learning Platform*. [Online]. Available: https://www.tensorflow.org/

[43] F. Chollet. (May 2015). *Keras*. [Online]. Available: https://keras.io

[44] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[45] A. Kent, M. M. Berry, F. U. Luehrs, and J. W. Perry, "Machine literature searching VIII. Operational criteria for designing information retrieval systems," *Amer. Document.*, vol. 6, no. 2, pp. 93–101, Apr. 1955.

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 Information Technology Convergence Lab, Department of Electronic Engineering, Myongji University (MJU), South Korea. Her research interests include statistics, data analysis, pattern recognition, computer vision, image processing, and deep learning techniques.

HA-ANH PHO received the bachelor's degree in biomedical engineering from the Hanoi University of Science and Technology (HUST), Hanoi, Vietnam, in June 2019. She is currently pursuing the master's degree with the IT Convergence Signal Processing Lab, Department of Electronic Engineering, Myongji University (MJU), South Korea. Her research interests include signal processing techniques with machine learning algorithms and applications of radar systems.

BYEONGHO LEE (Associate Member, IEEE) received the B.S. degree in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 2015, where he is currently pursuing the Ph.D. degree. His research interests include radar signal processing techniques, wireless indoor localization algorithms, and deep learning applications.

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), South Korea. From 2011 to 2013, he was an Assistant Professor with the Division of Marine 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 with the Faculty of the Korea National University of Transportation. His general research interests include communication systems, fault diagnosis, and digital signal processing. He is particularly interested in artificial intelligence for communications, radar systems, and smart grid.