


Letter

One-Shot Learning for Partial Discharge Diagnosis Using Ultra-High-Frequency Sensor in Gas-Insulated Switchgear

Vo-Nguyen Tuyet-Doan ¹, The-Duong Do ¹, Ngoc-Diem Tran-Thi ¹, Young-Woo Youn ² and Yong-Hwa Kim ^{1,*} 

¹ Department of Electronic Engineering, Myongji University, Yongin 17058, Korea; tuyetdoan201096@gmail.com (V.-N.T.-D.); theduong@mju.ac.kr (T.-D.D.); ngocdiem15051996@gmail.com (N.-D.T.-T.)

² HVDC Research Division, Korea Electrotechnology Research Institute (KERI), Changwon 51543, Korea; ywyoun@keri.re.kr

* Correspondence: yongkim@mju.ac.kr; Tel.: +82-31-330-6370

Received: 29 June 2020; Accepted: 25 September 2020; Published: 28 September 2020



Abstract: In recent years, deep learning has been successfully used in order to classify partial discharges (PDs) for assessing the condition of insulation systems in different electrical equipment. However, fault diagnosis using deep learning is still challenging, as it requires a large amount of training data, which is difficult and expensive to obtain in the real world. This paper proposes a novel one-shot learning method for fault diagnosis using a small dataset of phase-resolved PDs (PRPDs) in a gas-insulated switchgear (GIS). The proposed method is based on a Siamese network framework, which employs a distance metric function for predicting sample pairs from the same PRPD class or different PRPD classes. Experimental results over the small PRPD dataset that was obtained from an ultra-high-frequency sensor in the GIS show that the proposed method achieves outstanding performance for PRPD fault diagnosis as compared with the previous methods.

Keywords: ultra-high-frequency (UHF) sensor; gas-insulated switchgear (GIS); partial discharges (PDs); fault diagnosis; one-shot learning; Siamese network

1. Introduction

Power systems are being increasingly employed due to the increase in the demand for electricity, and their stability is important for stable operation of the power grid [1]. A gas-insulated switchgear (GIS) applied to substations is a major protection device for power facilities. GIS protects the power system by blocking excessive current quickly in the case of failure as well as normal opening and closing. When a failure occurs in a GIS, the impact of the accident is severe; hence, recovery takes a considerable amount of time and the power outage time increases. Various faults that cause an insulation breakdown of a GIS can be detected while using partial discharges (PDs) before insulation breakdown [2]. Therefore, detecting PDs in a GIS is necessary to ensure the safety and reliability of grid assets [3].

Electrical, mechanical, and chemical methods have been used in order to detect PDs in GISs [4–7]. Dissolved-gas analysis was employed in the chemical methods and ultra-high-frequency (UHF) sensors were used in the electromagnetic methods to detect PDs for on-site PD monitoring [8,9]. UHF sensors have the advantages of external interference immunity and high-sensitivity detection [10,11]. In this study, a UHF sensor was utilized for the PD measurement system [12].

Time-resolved PD (TRPD) and phase-resolved PD (PRPD) were used to investigate the PD characteristics in a GIS [13–15]. The TRPD-based methods analyze the PD pulses while

using time-domain, frequency-domain, and both time- and frequency-domain features [16–18]. The PRPD-based methods analyze the phase-amplitude-number (ϕ - q - n) of the PRPDs, where ϕ is the phase angle, q is the amplitude, and n is the number of PD occurrences [19]. The defect types were identified by analyzing the number of PD pulses, and the maximum amplitude or average amplitude in each phase [20]. Using the features from the PRPDs, machine-learning-based classifiers, such as support vector machines (SVMs) [21], decision trees [22], and neural networks [23,24], were proposed for PD classification.

Deep neural networks, which combine feature extraction and classification, have achieved promising results in several application areas, such as computer vision, natural language processing, speech recognition, and text classification [25]. Deep learning models, such as convolutional neural networks (CNNs) [26], recurrent neural networks (RNNs) [27], and self-attention [28], have been employed and state-of-the-art results have been obtained, in order to improve the performance for fault diagnosis while using a PRPD in a GIS. [26] proposed a CNN to learn the local response from the temporal or spatial signals of a PRPD. [27] proposed an RNN model with a long short-term memory to process sequential PRPD data. To overcome parallel computation and enable the capture of the interactions among PRPDs, [28] used a self-attention neural network to assess the possibility of focusing on important information and simultaneous computation from PRPD input. However, most of the existing deep-learning-based fault diagnosis methods require a large dataset for training and validation [26–28]. However, it is difficult to obtain a large dataset for fault diagnosis in the real world [27,28].

In this paper, we propose a new one-shot learning model for fault diagnosis with a small number of PRPDs while using Siamese neural networks that have the advantage of reducing the parameters to train and avoiding the problem of over-fitting [29]. One-shot learning and few-shot learning can learn features when only a few labeled samples are provided [30]. The proposed method uses a Siamese structure consisting of two identical CNNs and a distance metric function [31]. The two CNNs share the same parameters and map the PRPDs into a suitable embedding space. Subsequently, the distance metric function calculates the distance between the two CNN outputs. During the training phase, two PRPDs from the same class or different classes are paired and the paired sample is processed through the Siamese network to train the model for binary classification. The proposed model detects faults in the GIS while using the test PRPD pair and the support set from the training data. The experimental results with a small number of PRPDs show that the proposed one-shot learning model demonstrates a better performance for the PRPD classification than a CNN and linear SVM. The main contributions of this paper are summarized, as follows:

- One-shot learning is introduced for the first time to classify the PRPDs in a GIS. This method offers the advantages of a high classification accuracy while requiring a small amount of data compared with a linear SVM and CNN [30]. The proposed model uses pairs of samples of the same class or different classes during the training phase and recognizes the test sample with a single training sample for each class.
- The proposed model uses a distance metric function to map the PRPDs into a suitable embedding space and predicts the test PRPD class conditioned on the distance, which improves the classification performance as compared with that of the CNN [30].
- The proposed model is verified through PRPD and on-site noise measurements using a UHF sensor. The proposed model achieves a classification accuracy of 98.65% for four types of faults and noise in the GIS.

The remainder of this paper is organized, as follows: the PRPD and noise measurements in the GIS are introduced in Section 2. Section 3 presents the architecture of the proposed one-shot learning model. We compare the performance of the proposed method with that of other conventional methods in Section 4. Finally, we conclude this paper in Section 5.

2. Prpd and Noise Measurements

In this section, we present the PRPD and noise measurements that were obtained using a PD monitoring system for the GIS [27,28]. Figure 1 shows a block diagram of the PD monitoring system, consisting of a GIS, external UHF sensor, amplifier, a peak detector, and a data acquisition system (DAS). A cavity-backed patch antenna is used as the external UHF sensor, the amplifier has a gain of 45 dB, and the operating bandwidth is between 500 MHz and 1.5 GHz [27,28]. The peak detector is used to capture maximum values of UHF PD pulses [32]. After the peak detector, the DAS uses an analog-to-digital converter (ADC) with $1024 \times f_m$ samples per second, where $f_m = 60$ Hz is the power frequency. Subsequently, the maximum value is captured at every 8 samples in the DAS and $P = 128$ samples in each power cycle are used for PRPD measurements. The DAS uses an eight-bit analog-to-digital converter (ADC) with $P \times f_m$ samples per second, where $P = 128$ is the number of data points in each power cycle and $f_m = 60$ Hz is the power frequency. The measured signal at the p -th data point for the m -th power cycle is defined as $x(m, p) \in \{0, 1, \dots, 255\}$. Subsequently, the measured signal for $M = 3600$ power cycles is defined in a matrix form, as

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

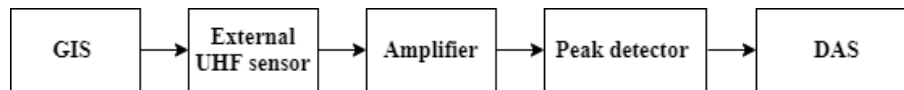


Figure 1. Block diagram of the partial discharge (PD) monitoring system.

2.1. Prpd Measurements in Gis

For the PRPDs in the GIS, we investigate four types of faults, namely, protruding electrodes, floating electrodes, void defects, and free particles, while using artificial cells [27,28]. Figure 2 shows four artificial cells used to simulate the defects in GISs [27,28], where an artificial cell is built for each fault. This is because failure in GIS occurs very rarely in on-site environments. For corona, the artificial cell simulated with a sharp protrusion fixed on an electrode caused a local electric field enhancement through a needle, with a tip radius of 10 μm and a diameter of 1 mm. The distance between the needle and the ground electrode was 10 mm, and the test voltage was 11 kV, as displayed in Figure 2a. To simulate an unconnected cell with a test voltage of 10 kV, Figure 2b shows the cell of a fabricated floating electrode (with the distances of 10 mm between the high-voltage (HV) and middle electrodes and 1 mm between the middle and ground electrodes). Small voids between the epoxy disc and the upper electrode were formed to simulate the artificial void discharge, at the test voltage of 8 kV, as shown in Figure 2c. Figure 2d simulates the free particle discharge with a test voltage of 10 kV. A small sphere with a diameter of 1 mm was placed on a concave ground electrode and the HV electrode was attached to a sphere of diameter 45 mm (fixed at 10 mm from the ground electrode). The artificial cells of the four faults were filled with 0.2 MPa of sulfur hexafluoride (SF_6) gas and the experiments under each failure are completed at one time.

Figure 3 shows the PRPDs for the four types of faults in three-dimensional (3D) and two-dimensional (2D) representations, where the amplitudes of the PRPDs for $M = 3600$ power cycles are accumulated to generate the 2D PRPD patterns, and the number of PD events per $M = 3600$ power cycles is represented by different colors. The measured signals for the corona fault exist on the negative half-cycle of the AC sine wave (from 180° to 360°). At approximately 45° of the cycle, the signal is low and close to zero. The floating PDs are presented during the first half of the positive and negative cycles of the AC sine wave with an extremely high signal density. The measured signals

for the void fault occur in the 45° – 90° and 180° – 270° phases with low amplitude. For the particle fault, a PD signal of relatively high amplitude is distributed across all of the phases.

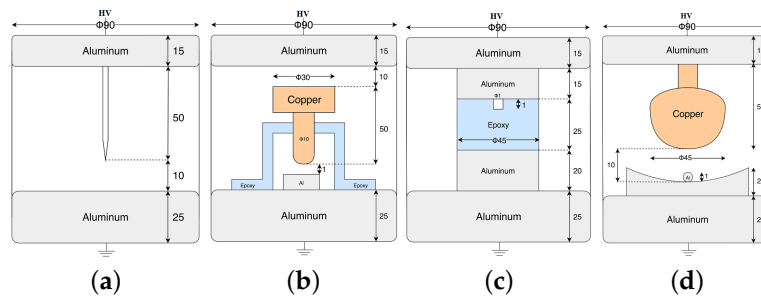


Figure 2. Artificial cells for the simulated (a) corona, (b) floating, (c) void, and (d) particle PDs.

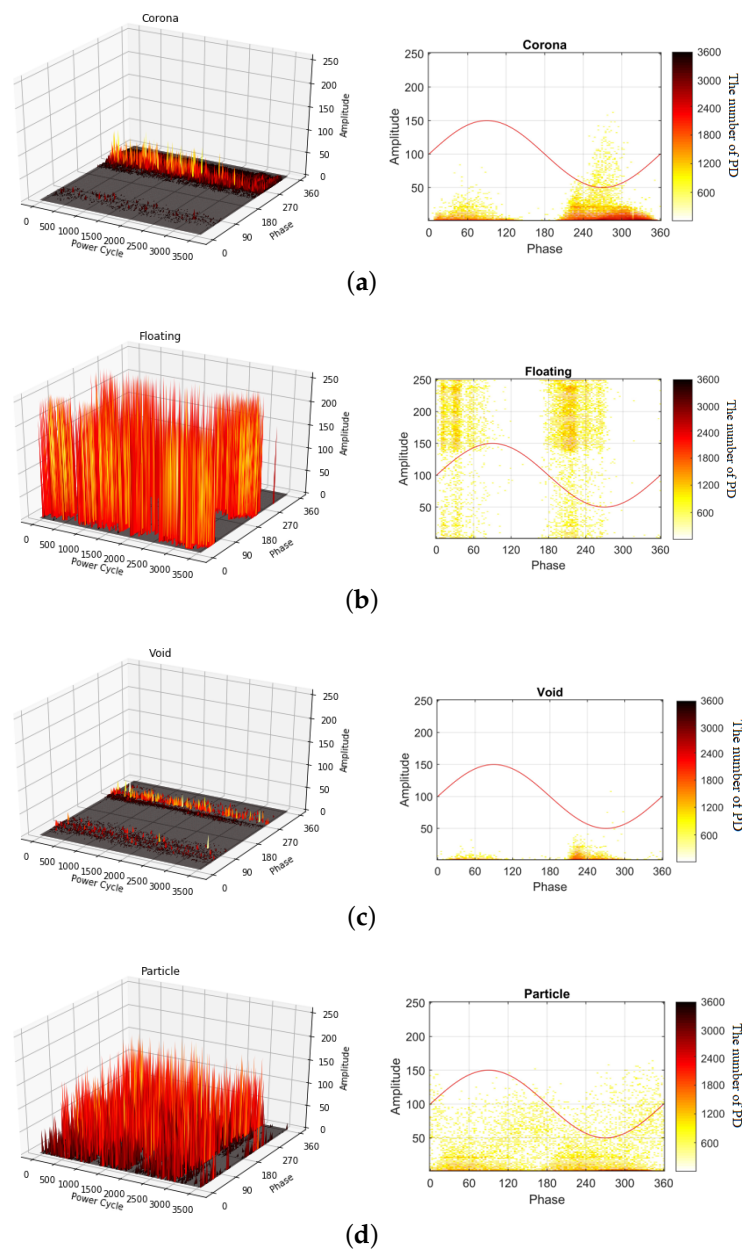


Figure 3. Examples of PRPDs in 3D and 2D representations: (a) corona, (b) floating, (c) void, and (d) particle faults.

Figure 4 shows changes in discharge characteristics with the duration of PRPD measurements. There are similar patterns between the beginning and the end of PRPDs since PRPD measurements were performed for tens of minutes.

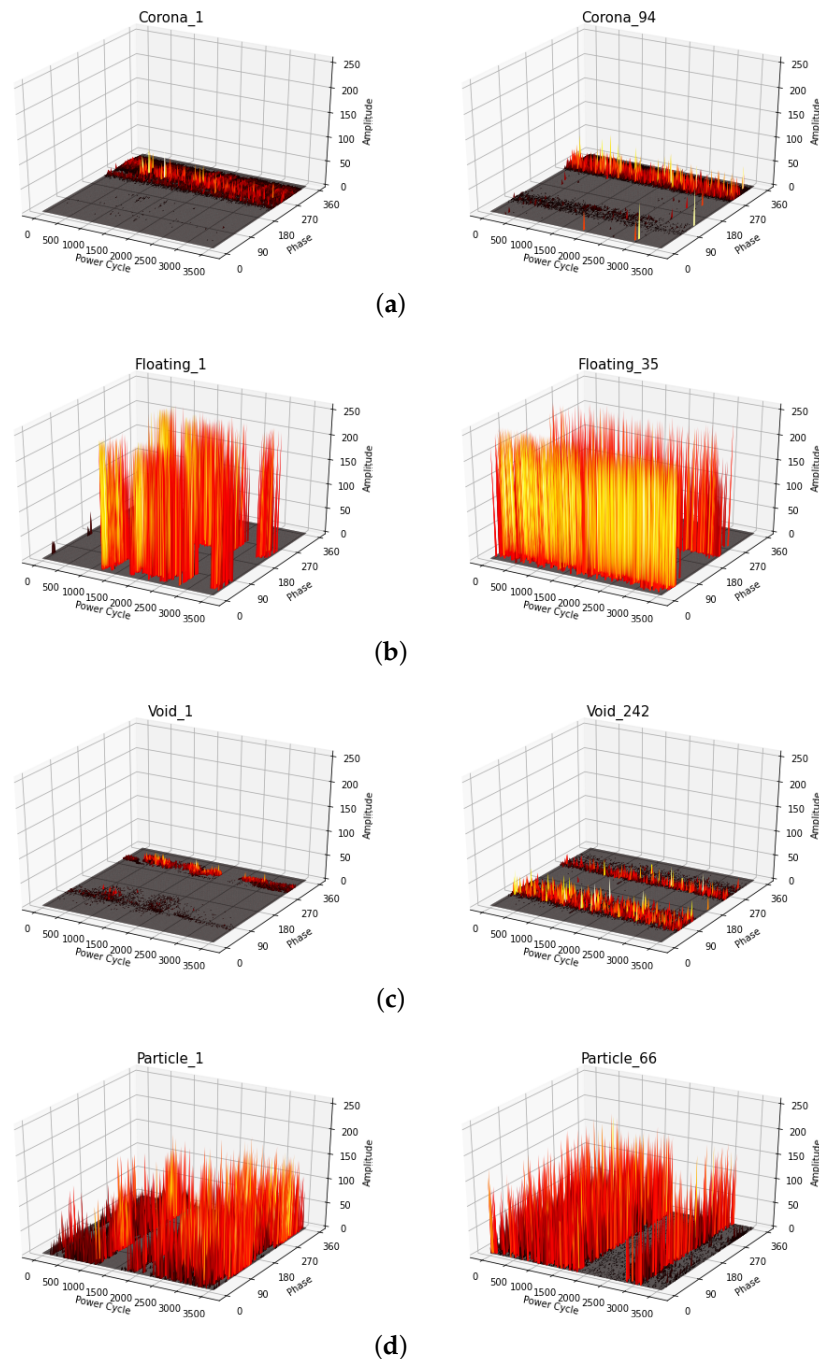


Figure 4. Discharge characteristics for the beginning (left) and the end (right) of PRPD measurements: (a) corona, (b) floating, (c) void, and (d) particle faults.

2.2. On-Site Noise Measurements

On-site noise was measured for 267 min. in a field in Korea [28]. Figure 5 shows the on-site noise measurements in the GIS, where the external UHF sensor is located on the outside of the spacer in the GIS. Figure 6 shows an example of the on-site noise signal measurement in 3D and 2D representations.

External noise signals appear across all of the phases in each power cycle and they are smaller than the PRPD signals in the GIS. In this paper, we regard on-site noise as a normal state for classification.



Figure 5. On-site noise measurements.

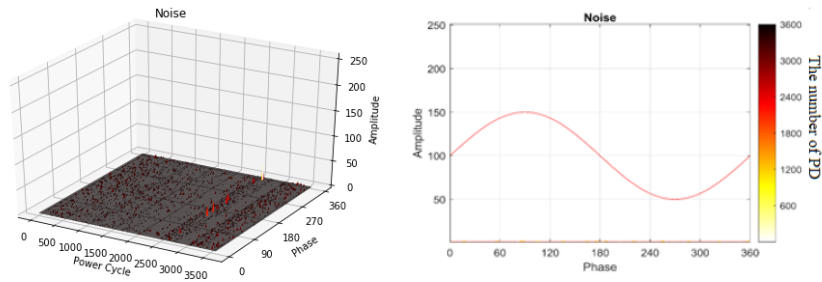


Figure 6. Example of on-site noise measurements in three-dimensional (3D) and two-dimensional (2D) representations.

3. Proposed Method

The proposed model for PRPD fault diagnosis is based on the architecture of one-shot learning [33]. Figure 7 shows the architecture of the proposed one-shot learning model, where the dataset is divided into three parts, namely, a training set \mathcal{T} , test set $\hat{\mathcal{T}}$, and support set \mathcal{S} . The training set \mathcal{T} is the collection of sample pairs from the same class or different classes. In the test phase, we denote a scenario with the support set \mathcal{S} consisting of K labels with N samples per class as K -way N -shot classification. Here, we consider that one PRPD ($N = 1$) is provided for each class as a sample for the support set \mathcal{S} , i.e., K -way one-shot classification for PRPD fault diagnosis.

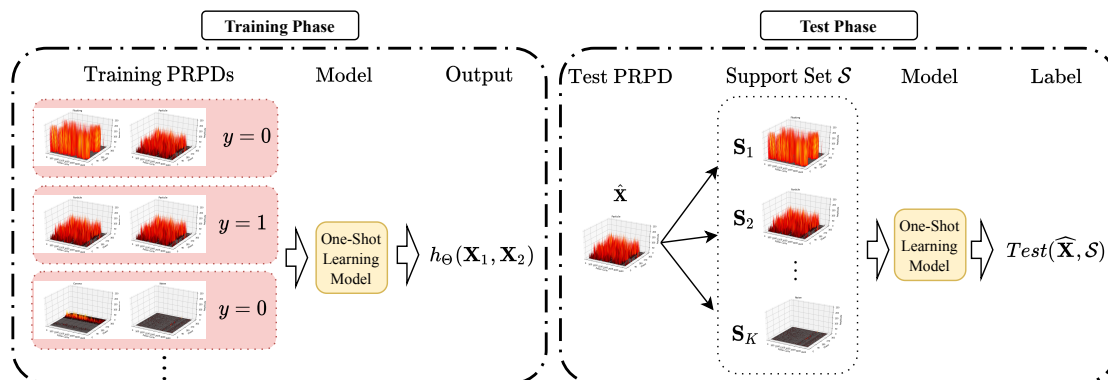


Figure 7. Architecture of the proposed one-shot learning model.

3.1. One-Shot Learning Model

Figure 8 shows the proposed one-shot learning model for PRPD fault diagnosis, which is based on a Siamese neural network [31]. The Siamese neural network is composed of an input layer, two identical CNNs, a distance metric, and an output layer.

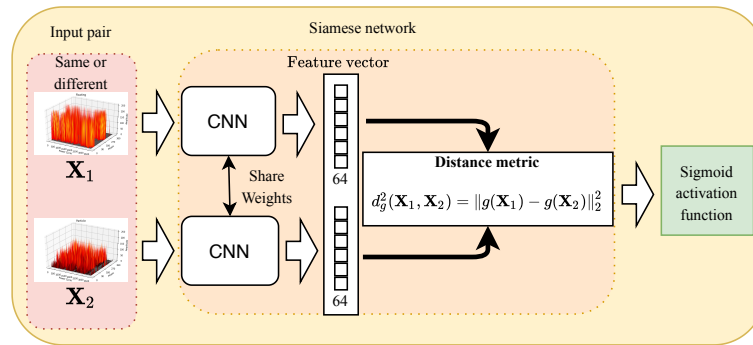


Figure 8. One-shot learning model using a Siamese network.

In the input layer, we denote $\mathbf{T} = (\mathbf{X}_1, \mathbf{X}_2) \in \mathcal{T}$ as an input pair of the same fault class or different fault classes, where \mathbf{X}_1 and \mathbf{X}_2 are the two PRPD signals in (1), respectively.

The same parameters and weights are used for the two identical CNNs, which are connected by a distance metric. Each CNN is composed of convolutional, max-pooling, dropout, flatten, and fully connected layers. The convolutional layers with multiple filters use the rectified linear unit (ReLU) activation function to improve the speed of the backpropagation computation and reduce the gradient-disappearance probability [34]. The max-pooling layer is used in order to reduce the number of computations and optimize the calculation space. Furthermore, dropout is added for regularization [35] and batch normalization is utilized in order to speed up learning by normalizing the input of the convolutional layers [36]. Flatten is used to transform a two-dimensional matrix into a vector. The distance metric is calculated based on the L2 norm as

$$d_g^2(\mathbf{X}_1, \mathbf{X}_2) = \|g(\mathbf{X}_1) - g(\mathbf{X}_2)\|_2^2, \quad (2)$$

where $g(\mathbf{X}_1)$ and $g(\mathbf{X}_2)$ are the feature vectors that are extracted by the CNN from \mathbf{X}_1 and \mathbf{X}_2 , respectively, and $g(\cdot)$ is the CNN.

In the proposed one-shot learning model, the output represents the probability of showing the similarity or dissimilarity between two input samples and it is related to the distance metric, as follows:

$$h_{\Theta}(\mathbf{X}_1, \mathbf{X}_2) = \sigma(\alpha d_g^2(\mathbf{X}_1, \mathbf{X}_2) + b), \quad (3)$$

where $\sigma(\cdot)$ is the sigmoid function, α is a weight value, and b is a bias. In addition, Θ is denoted as a vector containing every parameter in the model that should be determined.

3.2. Network Optimization

Figure 7 shows the one-shot learning training for PRPD fault diagnosis. For the training phase, we denote y as the label that corresponds to $\mathbf{T} = (\mathbf{X}_1, \mathbf{X}_2) \in \mathcal{T}$, where $y = 1$ if both \mathbf{X}_1 and \mathbf{X}_2 are from the same class, and $y = 0$, otherwise. From the PRPD measurement data, we create the set of input and output pairs (\mathbf{T}, y) to train and verify the proposed one-shot learning model. The parameters of the proposed model were learnt through the mini-batch \mathcal{B} in order to minimize the following loss function:

$$J(\Theta) = \frac{1}{|\mathcal{B}|} \sum_{b \in \mathcal{B}} \text{Loss}(b) + \frac{\lambda}{2|\mathcal{B}|} \|\Theta\|_2^2, \quad (4)$$

where $|\cdot|$ is the number of elements in a set and λ is the parameter for the L2-norm regularization. In (4), the loss for the b -th input and output pair is calculated based on cross entropy, as

$$Loss(b) = -y^{(b)} \log h_{\Theta}(\mathbf{X}_1^{(b)}, \mathbf{X}_2^{(b)}) - (1 - y^{(b)}) \log(1 - h_{\Theta}(\mathbf{X}_1^{(b)}, \mathbf{X}_2^{(b)})), \quad (5)$$

where the superscript (b) is used to indicate the index of the b -th input and output pair for the mini-batch \mathcal{B} .

Gradient descent is used to optimize neural networks and to alter the learning rate adaptively for minimizing the loss function. Numerous variations of the gradient descent method have been studied in previous studies e.g., AdaGrad, AdaDelta, Nesterov momentum into the ADAM, and Adam optimizer [37–40]. We select Adam as the optimizer with an initial learning rate of 6×10^{-5} [40].

In the testing phase, as shown in Figure 7, the support set $\mathcal{S} = \{\mathbf{S}_1, \dots, \mathbf{S}_K\}$ contains K PRPD training samples, where $\mathcal{S} \subset \mathcal{T}$ and each sample in the support set corresponds to each class. Subsequently, the test sample $\hat{\mathbf{X}} \in \hat{\mathcal{T}}$ and each element \mathbf{S}_k in the support set \mathcal{S} are entered into the Siamese network in order to calculate the similarity between the two inputs, where $k = 1, \dots, K$. The label for the test sample $\hat{\mathbf{X}} \in \hat{\mathcal{T}}$ is determined based on the greatest similarity as

$$Test(\hat{\mathbf{X}}, \mathcal{S}) = \arg \max_k (h_{\Theta}(\hat{\mathbf{X}}, \mathbf{S}_k)), \quad (6)$$

where the test set $\hat{\mathcal{T}}$ and the support set \mathcal{S} share the same label space. Finally, the accuracy of the proposed method is computed as

$$Accuracy = \frac{\text{Number of } Test(\hat{\mathbf{X}}, \mathcal{S}) \text{ is correctly classified}}{|\hat{\mathcal{T}}|} \times 100\%. \quad (7)$$

4. Experiment Results

We conducted PRPD experiments and noise measurements in the GIS in order to clarify the results of the one-shot learning algorithm for PRPD fault diagnosis. Table 1 shows the number of experiments for each fault, where the four PRPD faults, namely, corona, floating, particle, and void faults and noise, are coded by 0, 1, 2, 3, and 4, respectively. Each fault type of the PRPD signals and noise signals contains $M = 3600$ power cycles and each power cycle has $P = 128$ data points for one experiment.

Table 1. Experimental dataset.

Fault Types	Corona (0)	Floating (1)	Particle (2)	Void (3)	Noise (4)
Number of experiments	94	35	66	242	298

In our experiments, we use 81% samples of the data as the training set \mathcal{T} , 9% for the validation set \mathcal{V} , and the remaining as the test set $\hat{\mathcal{T}}$, where $|\mathcal{T}| = 594$, $|\mathcal{V}| = 67$, and $|\hat{\mathcal{T}}| = 74$, all of these sets are separate that have not each other appearing in its other stages. This has the advantage of demonstrating that the trained model does not have overfitting problems. We conducted extensive experiments to obtain hyperparameters for the different parameters used to tune our model. Some hyperparameters, such as the batch size, number of epochs, number of layers, kernel size, and number of kernels, were optimized. Table 2 illustrates the details of the CNN model in the proposed one-shot learning model. Moreover, we repeated the training process 10 times during our experiment in order to deal with the random initialization of the initial values for the training. The accuracy of the results was then averaged to confirm the validity of our model. All of the experiments were implemented based on TensorFlow [41] and Keras [42].

Table 3 shows the accuracy of the proposed one-shot learning model as compared with that of a linear SVM and CNN. The SVM is a well-known simple machine learning algorithm and the

CNN model uses the same structure as the Siamese network and a softmax function for multi-class classification. In our experiments, the linear SVM uses a feature vector for the maximum values at each phase to classify the faults in the GIS, while using the parameter $C = 1.0$. The proposed one-shot learning model achieved a classification accuracy of 98.65% and it showed a performance improvement of 2.7% over the CNN. This is because the distance metric of the proposed one-shot learning model uses two identical CNNs to map the PRPDs into a suitable embedding space and reduces the intra-class variation to avoid misclassification. In addition, the proposed one-shot learning model and CNN showed better performances than the SVM. The SVM requires appropriate feature extraction to train the classifier. However, the proposed one-shot learning model achieved a high classification accuracy while using a CNN structure without the feature extraction stage. For the corona fault, the classification accuracy was 100% for SVM, CNN, and the proposed one-shot learning model. The SVM showed the lowest performances of 25% and 28.57% for the floating and particle faults, respectively. This is because small training samples were provided for the floating and particle faults. The proposed one-shot learning model could classify each type of fault in the GIS and showed one error case in the void fault.

Table 2. Details of the convolutional neural networks (CNN) in the proposed one-shot learning model.

No.	Layer Type	Kernel Size/Stride	Kernel Number	Output Size	Padding
1	Convolution 1	$16 \times 16/4$	16	$900 \times 32 \times 16$	same
2	Max-Pooling 1	$2 \times 2/2$	-	$450 \times 16 \times 16$	valid
3	Batch Normalization 1	-	-	$450 \times 16 \times 16$	-
4	Drop Out 1	-	-	$450 \times 16 \times 16$	-
5	Convolution 2	$3 \times 3/1$	32	$450 \times 16 \times 32$	same
6	Max-Pooling 2	$2 \times 2/2$	-	$225 \times 8 \times 32$	valid
7	Batch Normalization 2	-	-	$225 \times 8 \times 32$	-
8	Drop Out 2	-	-	$225 \times 8 \times 32$	-
9	Convolution 3	$3 \times 3/1$	64	$225 \times 8 \times 64$	same
10	Max-Pooling 3	$2 \times 2/2$	-	$112 \times 4 \times 64$	valid
11	Batch Normalization 3	-	-	$112 \times 4 \times 64$	-
12	Drop Out 3	-	-	$112 \times 4 \times 64$	-
13	Convolution 4	$3 \times 3/1$	64	$112 \times 4 \times 64$	same
14	Max-Pooling 4	$2 \times 2/2$	-	$56 \times 2 \times 64$	valid
15	Batch Normalization 4	-	-	$56 \times 2 \times 64$	-
16	Drop Out 4	-	-	$56 \times 2 \times 64$	-
17	Flatten 1	-	-	7168×1	-
18	Dense 2	64	-	64×1	-

Table 3. Performance comparisons in terms of accuracy.

Fault Types	Overall (%)	Corona (%)	Floating (%)	Particle (%)	Void (%)	Noise (%)
Linear SVM	92.28	100	25	28.57	83.33	73.33
CNN	95.95	100	100	85.71	91.67	100
One-shot learning	98.65	100	100	100	95.83	100

We also conducted comparing the classification results with the balanced dataset by randomly selecting 35 samples for each PRPD fault from the whole experimental dataset. It can be seen that the performance of the one-shot learning and CNN are 94.44%, 88.89%, respectively, which are all dramatically higher than 72.22%, the performance of SVM, as shown in Table 4. Besides that, the proposed one-shot learning method performs better by about 5.55% higher in accuracy than the CNN method.

Figure 9 shows the confusion matrix for the one-shot learning model and the CNN. The proposed one-shot learning model has one error case for the void fault and the CNN has three errors (one error

for the particle fault and two errors for the void fault). All of the errors for the proposed one-shot learning model and CNN were misclassified as noise.

Table 4. Performance comparisons with balanced data case.

Fault Types	Overall (%)	Corona (%)	Floating (%)	Particle (%)	Void (%)	Noise (%)
Linear SVM	72.22	75	100	50	33.33	100
CNN	88.89	75	67.67	100	100	100
One-shot learning	94.44	75	100	100	100	100

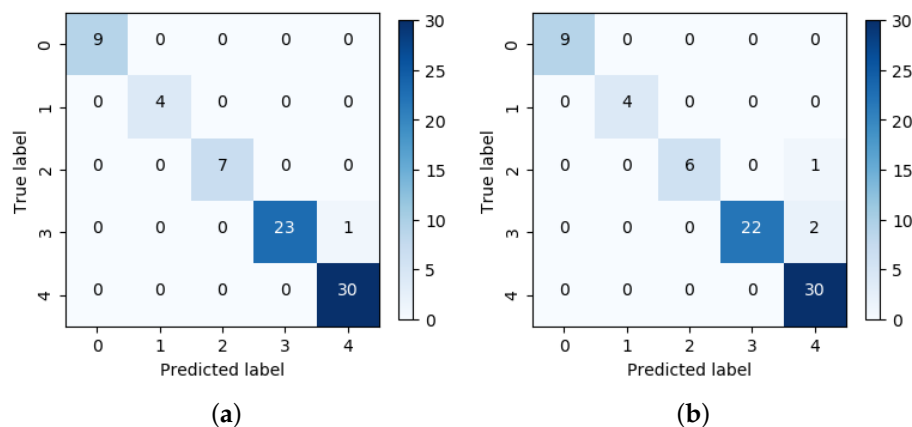


Figure 9. Confusion matrix with whole data case: (a) One-shot learning, (b) CNN.

The features of the proposed one-shot learning model were analyzed using t-distributed stochastic neighbor embedding (t-SNE) in order to understand the model learning better. Here, t-SNE reduces the dimensions of the data into 2D components with the maximum variation and visualizes them, such that similar features are transformed into nearby points. Figure 10 shows the t-SNE representations to visualize a set of inputs and their outputs of the one-shot learning model and the CNN, where the last fully connected layers $g(X)$ in the Siamese network and the CNN are used. Figure 10a shows that numerous fault data are very close to the noise and, hence, are difficult to classify accurately using the input PRPDs. Figure 10b shows that the features of each fault for the one-shot learning model are separately distributed. As shown in Figure 10c, the particle and void faults slightly overlap with the noise sample using the CNN and, hence, the CNN has some errors in the particle and void faults.

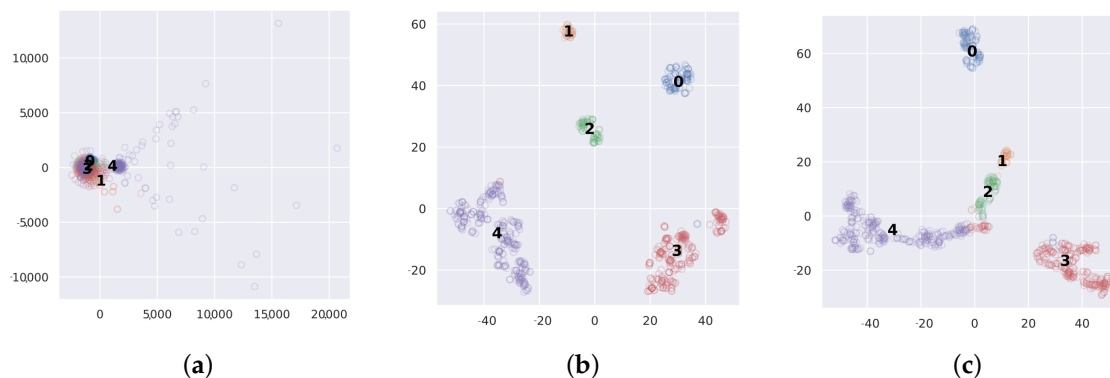


Figure 10. Visualization using t-SNE with whole data case: (a) Input, (b) One-shot learning, and (c) CNN.

5. Conclusions

In this study, we classified the PRPDs in a GIS while using a small amount of training data. Because fault diagnosis using deep learning will not fit well with small data, this paper addresses the challenge by adopting the Siamese network for one-shot learning. First, we used an external UHF sensor to acquire PRPD measurements from artificial cells and on-site noise in the GIS. Subsequently, we proposed a one-shot learning method based on a Siamese neural network framework. The proposed method extracts PRPD features from two identical CNNs, measures the distance between the PRPD features, and predicts whether their output pair is considered to be from the same class or different classes. Finally, we compare the experimental results for dataset cases: whole data and balanced data. The experimental results showed that the performance of the proposed method for PRPD classification with a small dataset was better than that of the SVM and CNN.

For future studies, we intend to design artificial cells of surface/creeping discharges on the surface of the GIS insulator for analyzing the PRPD patterns and conduct further verification of the proposed method for on-site fault data.

Author Contributions: Y.-H.K. conceived of the presented idea. V.-N.T.-D., T.-D.D. and N.-D.T.-T. developed the model and performed the computation. Y.-W.Y. verified the experimental setup and results. All authors discussed the results and contributed to the final manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was in part supported by Korea Electric Power Corporation [Grant number: R18XA01]; and in part 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 [No. 20179310100050].

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Riera-Guasp, M.; Antonino-Daviu, J.A.; Capolino, G.-A. Advances in Electrical Machine, Power Electronic, and Drive Condition Monitoring and Fault Detection: State of the Art. *IEEE Trans. Ind. Electron.* **2015**, *62*, 1746–1759. [\[CrossRef\]](#)
2. Lu, C.; Wenrong, S.; Chenzhao, F.; Kai, G. The Development of UHF Sensor for GIS Partial Discharge Measurement. In Proceedings of the 2019 IEEE 2nd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE), Shenyang, China, 22–24 November 2019; pp. 287–290.
3. Rao, M.M.; Kumar, M. Ultra-High Frequency (UHF) Based Partial Discharge Measurement in Gas Insulated Switchgear (GIS). In Proceedings of the 2019 IEEE International Conference on High Voltage Engineering and Technology (ICHVET), Hyderabad, India, 7–8 February 2019; pp. 1–5.
4. Phung, B.T.; Blackburn, T.R.; Liu, Z. Acoustic measurements of partial discharge signals. *J. Electr. Electron. Eng. Aust.* **2001**, *21*, 41.
5. Descouedres, A.; Hollenstein, C.; Demellayer, R.; Walder, G. Optical emission spectroscopy of electrical discharge machining plasma. *J. Mater. Process. Technol.* **2004**, *149*, 184–190. [\[CrossRef\]](#)
6. Biswas, S.; Koley, C.; Chatterjee, B.; Chakravorti, S. A methodology for identification and localization of partial discharge sources using optical sensors. *IEEE Trans. Dielectr. Electr. Insul.* **2012**, *19*, 18–28. [\[CrossRef\]](#)
7. Duval, M. A review of faults detectable by gas-in-oil analysis in transformers. *IEEE Electr. Insul. Mag.* **2002**, *18*, 8–17. [\[CrossRef\]](#)
8. Han, X.; Li, J.; Zhang, L.; Pang, P.; Shen, S. A Novel PD Detection Technique for Use in GIS Based on a Combination of UHF and Optical Sensors. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 2890–2897. [\[CrossRef\]](#)
9. IEC-62478. *High-Voltage Test Techniques—Measurement of Partial Discharges by Electromagnetic and Acoustic Methods. Proposed Horizontal Standard*, 1st ed.; International Electrotechnical Commission (IEC): Geneva, Switzerland, 2016.
10. Hoshino, T.; Koyama, H.; Maruyama, S.; Hanai, M. Comparison of sensitivity between UHF method and IEC 60270 for onsite calibration in various GIS. *IEEE Trans. Power Deliv.* **2006**, *21*, 1948–1953. [\[CrossRef\]](#)
11. Tenbohlen, S.; Denissov, D.; Hoek, S.; Markalous, S.M. Partial discharge measurement in the ultra high frequency (UHF) range. *IEEE Trans. Dielectr. Electr. Insul.* **2008**, *15*, 1544–1552. [\[CrossRef\]](#)

12. Suprianto, I.; Khayam, U.; Nishigouchi, K.; Kamarol, M.; Kozako, M.; Hikita, M. UHF sensor optimization used for detecting partial discharge emitted electromagnetic wave in gas insulated switchgear. In Proceedings of the 2014 IEEE International Symposium on Electrical Insulating Materials, Niigata, Japan, 1–5 June 2014; pp. 229–232.
13. Raymond, W.J.K.; Illias, H.A.; Bakar, A.H.A.; Mokhlis, H. Partial discharge classifications: Review of recent progress. *Measurement* **2015**, *68*, 164–181. [\[CrossRef\]](#)
14. Mor, A.R.; Heredia, L.C.C.; Munoz, F.A. Estimation of charge, energy and polarity of noisy partial discharge pulses. *IEEE Trans. Dielect. Electr. Insul.* **2017**, *24*, 2511–2521. [\[CrossRef\]](#)
15. Mor, A.R.; Morshuis, P.H.F.; Smit, J.J. Comparison of charge estimation methods in partial discharge cable measurements. *IEEE Trans. Dielect. Electr. Insul.* **2015**, *22*, 657–664. [\[CrossRef\]](#)
16. Nair, R.P.; Vishwanath, S.B. Analysis of partial discharge sources in stator insulation system using variable excitation frequency. *IET Sci. Meas. Technol.* **2019**, *13*, 922–930. [\[CrossRef\]](#)
17. Chen, X.; Qian, Y.; Sheng, G.; Jiang, X. A time-domain characterization method for UHF partial discharge sensors. *IEEE Trans. Dielect. Electr. Insul.* **2017**, *24*, 110–119. [\[CrossRef\]](#)
18. Jahangir, H.; Akbari, A.; Werle, P.; Akbari, M.; Szczechowski, J. UHF characteristics of different types of PD sources in power transformers. In Proceedings of the 2017 IEEE Iranian Conference on Electrical Engineering (ICEE), Tehran, Iran, 2–4 May 2017; pp. 1242–1247.
19. Sahoo, N.C.; Salama, M.M.A.; Bartnikas, R. Trends in partial discharge pattern classification: A survey. *IEEE Trans. Dielect. Electr. Insul.* **2005**, *12*, 248–264. [\[CrossRef\]](#)
20. Karthikeyan, B.; Gopal, S.; Venkatesh, S. Partial discharge pattern classification using composite versions of probabilistic neural network inference engine. *Expert Syst. Appl.* **2008**, *34*, 1938–1947. [\[CrossRef\]](#)
21. Umamaheswari, R.; Sarathi, R. Identification of Partial Discharges in Gas-insulated Switchgear by Ultra-high-frequency Technique and Classification by Adopting Multi-class Support Vector Machines. *Electr. Power Components Syst.* **2011**, *39*, 1577–1595. [\[CrossRef\]](#)
22. Abdel-Galil, T.; Sharkawy, R.; Salama, M.; Bartnikas, R. Partial Discharge Pattern Classification Using the Fuzzy Decision Tree Approach. *IEEE Trans. Instrum. Meas.* **2005**, *54*, 2258–2263. [\[CrossRef\]](#)
23. Darabad, V.P.; Vakilian, M.; Blackburn, T.R.; Phung, B.T. An efficient PD data mining method for power transformer defect models using SOM technique. *Int. J. Electr. Power Energy Syst.* **2015**, *71*, 373–382. [\[CrossRef\]](#)
24. Abubakar Mas'ud, A.; Stewart, B.G.; McMeekin, S.G. Application of an ensemble neural network for classifying partial discharge patterns. *Electr. Power Syst. Res.* **2014**, *110*, 154–162. [\[CrossRef\]](#)
25. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#)
26. Li, G.; Wang, X.; Li, X.; Yang, A.; Rong, M. Partial Discharge Recognition with a Multi-Resolution Convolutional Neural Network. *Sensors* **2018**, *18*, 3512. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Nguyen, M.T.; Nguyen, V.H.; Yun, S.J.; Kim, Y.H. Recurrent Neural Network for Partial Discharge Diagnosis in Gas-Insulated Switchgear. *Energies* **2018**, *11*, 1202. [\[CrossRef\]](#)
28. Tuyet-Doan, V.-N.; Nguyen, T.-T.; Nguyen, M.-T.; Lee, J.-H.; Kim, Y.-H. Self-Attention Network for Partial-Discharge Diagnosis in Gas-Insulated Switchgear. *Energies* **2020**, *13*, 2102. [\[CrossRef\]](#)
29. Hsiao, S.-C.; Kao, D.-Y.; Liu, Z.-Y.; Tso, R. Malware Image Classification Using One-Shot Learning with Siamese Networks. *Procedia Comput. Sci.* **2019**, *159*, 1863–1871. [\[CrossRef\]](#)
30. Zhang, A.; Li, S.; Cui, Y.; Yang, W.; Dong, R.; Hu, J. Limited Data Rolling Bearing Fault Diagnosis with Few-Shot Learning. *IEEE Access* **2019**, *7*, 110895–110904. [\[CrossRef\]](#)
31. Koch, G.; Zemel, R.; Salakhutdinov, R. Siamese neural networks for one-shot image recognition. In Proceedings of the 2nd ICML Deep Learning Workshop, Lille, France, 10–11 July 2015.
32. Vedral, J.; Kříž, M. Signal Processing in Partial Discharge Measurement. *Metrol. Meas. Syst.* **2010**, *17*, 55–64. [\[CrossRef\]](#)
33. Wang, B.; Wang, D. Plant Leaves Classification: A Few-Shot Learning Method Based on Siamese Network. *IEEE Access* **2019**, *7*, 151754–151763. [\[CrossRef\]](#)
34. Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML'10), Haifa, Israel, 21–24 June 2010; Volume 27, pp. 807–814.
35. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.

36. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
37. Duchi, J.; Hazan, E.; Singer, Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *J. Mach. Learn. Res.* **2011**, *12*, 2121–2159.
38. Zeiler, M.D. ADADELTA: An Adaptive Learning Rate Method. *arXiv* **2012**, arXiv:1212.5701.
39. Dozat, T. Incorporating Nesterov Momentum into Adam. In Proceedings of the ICLR Workshop, Caribe Hilton, San Juan, Puerto Rico, 2–4 May 2016; pp. 2013–2016.
40. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference for Learning Representations, San Diego, CA, USA, 7–9 May 2015; pp. 807–814.
41. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.; Davis, A.; Dean, J.; Devin, M. TensorFlow : Large-Scale Machine Learning on Heterogeneous Distributed Systems. 2015. Available online: <https://www.tensorflow.org/> (accessed on 8 June 2020).
42. Keras-Team. 2015. Available online: <https://github.com/fchollet/keras> (accessed on 8 June 2020).



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).