

Article

Partial Discharge Online Detection for Long-Term Operational Sustainability of On-Site Low Voltage Distribution Network Using CNN Transfer Learning

Jinseok Kim ¹ and Ki-Il Kim ^{2,*} 

¹ Department of KDN Electric Power IT Research Institute, KEPCO KDN, Naju 58322, Korea; kim-668069@kdn.com

² Department of Computer Science and Engineering, Chungnam National University, Daejeon 34134, Korea

* Correspondence: kikim@cnu.ac.kr; Tel.: +82-42-821-6856

Abstract: Partial discharge (PD) detection studies aiming at the fault diagnosis for facilities and power cables in transmission networks have been conducted over the years. Recently, the deep learning models for PD detection have been used to diagnose the PD fault of facilities and cables. Most PD studies have been conducted in the field, such as gas-insulated switchgear (GIS) and power cables for high voltage transmission networks. There are few studies of PD fault detection for on-site low-voltage distribution networks. Additionally, there are few studies of PD detection algorithms for improving the accuracy of the deep learning models using small real PD data only. In this study, a PD online detection system and a model for long-term operational sustainability of on-site low voltage distribution networks are proposed using convolutional neural network (CNN) transfer-learning. The proposed PD online system makes it possible to acquire as many real PD data as possible through continuous monitoring of PD occurrence. The PD detection accuracy results showed that the proposed CNN transfer-learning models are more effective models for obtaining improved accuracy (97.4%) than benchmark models, such as CNN and support vector machine (SVM) using only small real PD data acquired from PD online detection system.

Keywords: partial discharge; online detection; distribution network; operational sustainability; transfer learning



Citation: Kim, J.; Kim, K.I. Partial Discharge Online Detection for Long-Term Operational Sustainability of On-Site Low Voltage Distribution Network Using CNN Transfer Learning. *Sustainability* **2021**, *13*, 4692. <https://doi.org/10.3390/su13094692>

Academic Editor: Tomonobu Senjyu

Received: 1 April 2021
Accepted: 21 April 2021
Published: 22 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Partial discharge (PD) is defined as a localized insulation breakdown between conductors based on IEC-60270 [1]. PD detection studies play a significant role in judging the insulation status of cable, transmission, and distribution systems to prevent major faults of power systems and enhance the reliability and the long-term operational sustainability of the electric power supply [2]. Currently, there are several issues regarding PD detection studies in the aspects of system sustainability for life cycle assessment and management.

First, most PD detection studies have been conducted in the field, such as gas-insulated switchgear (GIS) and power cables for high voltage transmission networks [2,3]. There are few studies of PD fault detection for low voltage distribution networks closest to the electricity customers.

There are many facilities and cables for low-voltage distribution networks in South Korea, and most of them are more than 30 years old. Old facilities and cables can easily cause major electricity faults. This is the reason why continuous PD diagnosis is crucial. To diagnose many distribution facilities and cables continuously and accurately in all areas of Korea, we need many PD experts for PD diagnosis. However, excessive expert diagnosis to identify the status of many distribution facilities or cables can cause human errors for PD detection and increase the time and maintenance costs [4]. This is the reason

why PD studies regarding an automated diagnosis for low voltage distribution networks are required.

Another reason why PD studies for low voltage distribution networks are important is the physical and electrical differences between distribution and transmission networks.

Physically, the distribution lines cover shorter distances and provide electricity to residential consumers through complicated multiple routes compared to high voltage transmission lines.

Unlike high-voltage (154 kV and 345 kV) transmission lines, the voltage standards for distribution networks in South Korea are 220 V in a single phase at 60 Hz or 220/380 V in three-phase at 60 Hz. As shown in [5,6], the equipment in the transmission network handles high voltage (154 kV and 345 kV) compared with the equipment in the distribution network. The authors of [7] reported that the variations in voltage magnitude and phase angle are different between lower and higher voltages. Thus, the PD study between the distribution network and the transmission network are different based on the fact that the PD detection is basically performed based on the voltage and phase.

Considering these points, the physical and electrical differences could not make it possible to apply PD studies for the high voltage transmission network to PD studies for the low voltage distribution network.

In addition, there are few studies of PD detection models for improving the accuracy of the deep learning models using only small real PD data. Recently, PD detection has been conducted using deep learning models. Many deep learning models tend to use artificial PD data generated from the laboratory for developing the model due to insufficient real PD data. It is rather difficult to acquire sufficient real PD data since the real PD occurrence is a nonperiodic and unpredictable phenomenon. To improve the PD detection reliability of the deep learning model, it is essential to develop a deep learning model with as many real PD data as possible. The above-mentioned problems for long-term operational sustainability are summarized as follows:

- Few studies of PD fault detection in low voltage distribution network;
- Excessive expert diagnosis to obtain high possibilities to cause human errors of PD detection and increase the time and cost maintenance;
- Difference in physical and electrical properties between distribution and transmission networks;
- Development of deep learning models to use artificial PD data generated from the laboratory for developing the model due to insufficient real PD data.

In this study, the PD online detection system that combines data collection and transmission unit (DCTU), high-frequency current transformer (HFCT) sensor, and operating server for long-term operational sustainability of on-site low voltage distribution network is proposed to overcome these issues. The PD online detection system makes it possible to acquire as many real PD data as possible through continuous monitoring of PD occurrence. In addition, convolutional neural network (CNN) transfer-learning models are proposed as a PD detection algorithm. It is well-known that transfer-learning models are effective methods for improving accuracy performance using small data. To verify the effectiveness of the transfer-learning model of PD detection for a low distribution network, the study on CNN transfer-learning models for improving PD detection accuracy is conducted using only a small real PD dataset. The main contributions of this study are summarized as follows:

- A field study of PD online detection for long-term operational sustainability of on-site low voltage distribution network;
- An automated PD online detection system for acquiring as many real PD data as possible through the continuous monitoring of PD occurrence;
- Verification of the effectiveness of the CNN transfer-learning models developed using only a small real PD dataset;

- Improved PD detection accuracy of the proposed CNN transfer-learning models compared with benchmark models, such as CNN and support vector machine (SVM) models.

The remainder of this paper is organized as follows. Section 2 introduces various studies of PD fault detection. Section 3 describes PD online detection system architecture and PD detection algorithm based on CNN transfer-learning models. Section 4 presents PD detection accuracy results of the proposed and benchmark models. Section 5 summarizes the effectiveness of the proposed method. Finally, Section 6 presents the conclusion.

2. Related Studies

The article [8] reviewed recent PD identification study results using several deep learning methods. First, most of them showed that the CNN model is widely used as a PD fault diagnosis model. Second, most models were trained by artificial PD data and few real PD samples. This can be a problem in on-site installations. Third, the online PD monitoring system is the main solution to diagnose PD faults of facilities and cables using deep learning algorithms. The PD deep learning models have better accuracy than machine learning models and show effective automated recognition. In [1], the autoencoder model was used as a PD identification method. The deep learning model classifier consists of a sparse autoencoder layer and softmax function using artificial PD defects acquired by PD current waveform probe and ultrahigh frequency (UHF) sensor. The proposed method achieved an accuracy of 99.7%. In [2], the authors presented a CNN model employed for PD fault diagnosis from high voltage cables. The studies were executed on artificial defects in the laboratory, and the proposed CNN model showed superior accuracy compared with traditional models: SVM and backpropagation neural network (BPNN). In [9], the machine learning ensemble method was used for PD localization. The regression tree, random forest, and bootstrap aggregating were applied using the wavelet packet transform (WPT). The proposed method showed improved performance (91%) compared with a benchmark regression tree model. In [10], a PD detection method was proposed for the power transformer using a UHF sensor. In [11], automatic online PD was diagnosed using CNN deep learning model. The proposed system employed PD monitoring data, ultrasonic sensing, and a CNN model. The article [12] investigated the ensemble method and long short-term memory (LSTM) deep learning for PD classification. The ensemble bagged decision trees, and LSTM deep learning showed accuracies of 95.5% and 98.3%, respectively. In [13], the CNN models' effectiveness was verified for PD analysis. The PD data were collected from actual in-service cables, and the data type was PD waveform in the time domain. The CNN models were basic CNN and CNN-recurrent neural network (CNN-RNN). The two CNN models achieved 95% accuracy compared with 87.92% accuracy of multilayered perceptron (MLP). In [14], one-dimensional (1D) CNN based on time domain was used for PD recognition. The 1D CNN was used for reducing the model complexity, and a time-domain waveform was extracted using a UHF sensor. Furthermore, 70% and 30% of samples were real and experimental datasets, respectively. The proposed model's accuracy is 88.9%. In [15], transient earth voltage (TEV) sensors and CNN models were used for PD fault detection of switchgear. The CNN model was trained and tested using 360 artificial and 16 real data. The article [16] investigated the CNN model with TEV, surface current sensor (SCS), and HFCT sensor. The input data were PD waveform in the time-domain. It has the highest accuracy of 99.7% for the HFCT sensor. In [17], the CNN model was used to diagnose PD fault for cable accessories with 800 data samples, and a 10-fold cross-validation method was used. The author reported that the recognition rate was 92.5% and the performance was improved compared with SVM and BPNN. In [18], one-shot learning was investigated for PD diagnosis using UHF sensor in GIS. The proposed model employed a distance metric function for PD fault classification. The model achieved 98.65% accuracy to classify faults and noise. In [19], three CNN subnetworks were used for PD recognition using UHF sensors as an input extraction, and an LSTM network was finally employed to classify PD faults using the output of CNN networks. It explained the

effectiveness of a unified framework for CNN and LSTM. In [20], sparse autoencoder (SAE) and deep belief networks (DBN) were used for cable PD identification. SAE was used for unsupervised pretraining, whereas DBN was used for supervised fine-tuning to detect PD faults. The dataset was acquired from PSCAD/EMTDC simulation, and 16,800 artificial samples in six categories were obtained from the simulation. In [21], CNN and LSTM models were incorporated for PD recognition using a UHF sensor. CNN extracted local spatial features, while LSTM was used to extract timing features. The extracted pattern was finally classified using a fully connected layer and softmax function. The article [22] investigated CNN models with several layers, max pooling, and dropout with PD images. The models were used and tested for high voltage insulation defects. It achieved the highest accuracy of 97.41%. The authors [3] proposed variational autoencoder (VAE) using a data matching method. Cosine distance was used to classify PD data. The model was developed using a laboratory PD fault dataset for GIS application. In [23], the combination of LSTM and self-attention was proposed for PD diagnosis in GIS. Four types of PD faults were used, and the augmentation process was used to increase training samples. The accuracy of the models was superior compared to the benchmark and LSTM-RNN models. In [4], a light-scale CNN model was used for PD recognition in GIS. Artificial defects were created in the experiment, and a UHF sensor was used to acquire PD faults. The light-scale model incorporated two convolution layers, two max pooling, and two fully connected layers. The overall accuracy of the model was 98.13%.

The above studies address various PD detection techniques using machine learning and deep learning models. As shown in Figure 1, the proposed system and model show noble features compared with the above PD studies as follows:

- The automated PD online detection system can provide as many real PD data as possible through the continuous monitoring of PD occurrence for on-site low voltage distribution networks.
- The proposed CNN transfer-learning models can obtain improved accuracy compared to benchmark models with only real PD data acquired from the PD detection system.

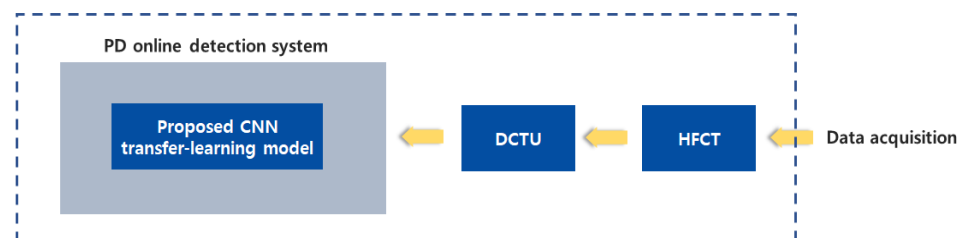


Figure 1. The scope and relations for the proposed method.

3. Proposed Method

In the low-voltage distribution network, there are two main electricity facilities: transformers, which convert high voltage (22.9 kV) to low voltage (220/380 V), and switches, which connect or disconnect the electricity path in the distribution network. Furthermore, there are electric manholes in the ground. Under the manholes, there are many electricity power cables. When these facilities and cables become old, they can cause PD faults. In reality, the PD events has been frequently reported from two electricity facilities (transformer and switches) and electricity power cables. Therefore, regarding the study, we focus on two main electricity facilities and electricity power cables in this paper. The proposed PD detection system can monitor these facilities and cables to identify PD faults continuously in real-time. The PD detection system has been set up at Daejeon in South Korea. The user interface of the operating server for the PD detection system can intuitively identify PD detection in real-time.

3.1. PD Online Detection System Architecture

The PD detection system architecture is described in Figure 2. As shown in Figure 2, the system architecture incorporates the HFCT sensor, DCTU, and operating server. The HFCT sensor detects the PD data, the DCTU collects the PD data acquired from the HFCT sensor, and then the DCTU transmits the PD data to the operating server using wireless hopping or power line communication (PLC) between DCTUs.

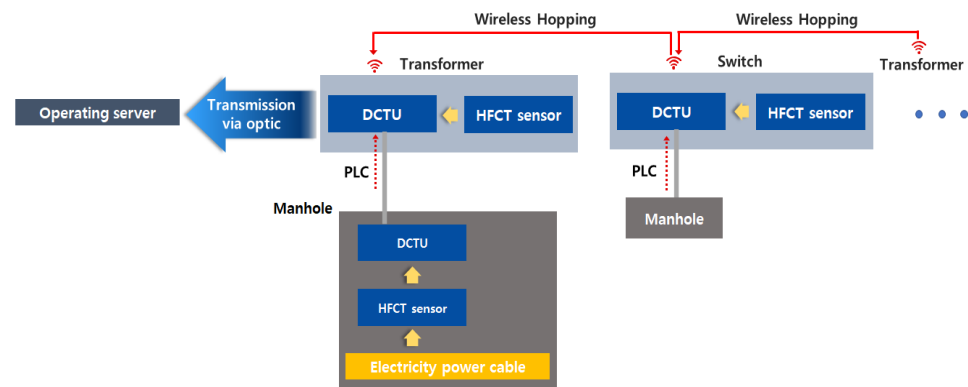


Figure 2. PD detection system architecture incorporating HFCT sensor, DCTU, and operating server.

First, the HFCT sensor, as a PD detection sensor, is used for the PD detection system. The HFCT sensor has been widely used for PD pulse detection [2,13] in transmission networks. The HFCT sensor can detect high-frequency PD pulse in the range from 50 kHz to 20 MHz. In this study, the HFCT sensor is used to detect PD pulses for low-voltage distribution networks. The technical specification of the HFCT sensor is shown in Table 1. The bandwidth of the HFCT sensor ranges from 50 kHz to 20 MHz. The load impedance is 50 Ω , and the sensitivity is 15 mV/mA. The sensor has a clamp type, which can easily be installed or removed on a cable. The HFCT sensor transmits the data to the DCTU using a coaxial cable with a BNC connector.

Table 1. Technical specification of the HFCT sensor for the low distribution network.

Information for HFCT	Specification
Bandwidth	50 kHz–20 MHz
Load impedance	50 Ω
Sensitivity	15 mV/mA
Type	Clamp
Connector	BNC

Figure 3 shows the PD detection methods using the HFCT sensor. The HFCT sensor is clamped on three phases of electricity cables in electric power distribution facilities. The HFCT sensor can detect the generated PD pulses among three electricity phases of the facility in the power distribution facilities. Moreover, the PD pulses generated from the electric power distribution facilities can be captured.

The DCTU collects the data acquired from the HFCT sensor and transmits the PD data to the operating server using wireless hopping or PLC communication between DCTUs. The functional block diagram for DCTU is shown in Figure 4. The DCTU consists of four parts: power, CPU, memory, and communication. The power module incorporates power current transformer (CT), surge protector, and voltage regulator. The power CT unit supplies power to activate the DCTU. The surge protector unit protects the DCTU against an inrush current. The voltage regulator maintains the voltage that each block (CPU, memory, and communication) can use. The memory stores the data. The ARM Cortex A9 is used as a CPU. The communication block employs Wi-Fi HaLow, optic, and PLC for data transmission.

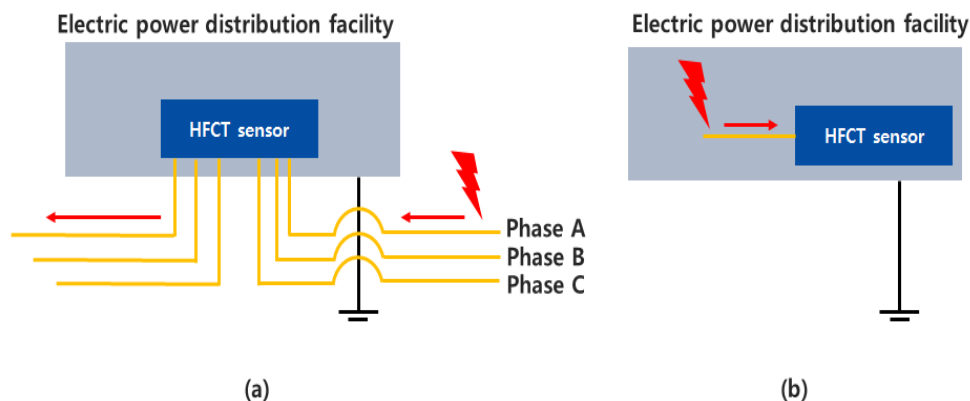


Figure 3. PD detection methods using HFCT sensor: (a) PD detection among three electricity phases and (b) PD detection within an electric power distribution facility.

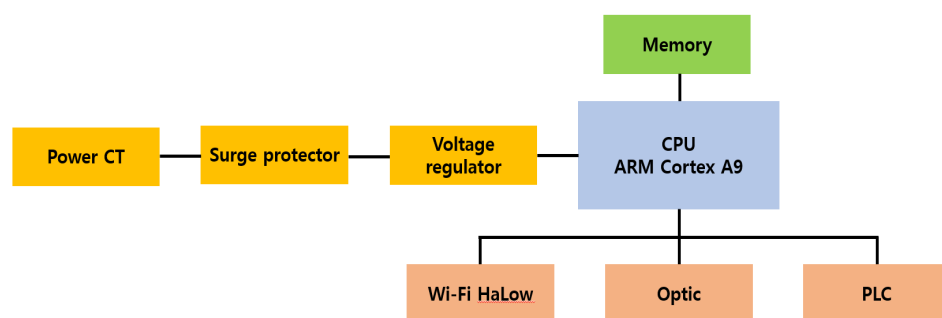


Figure 4. Functional block diagram for DCTU.

Figure 5 shows PLC, wireless hopping, and optic communication for data transmission. PLC is used to transmit the data generated from cables under the manhole. The wireless hopping method is used to transmit the collected data to the operating server between DCTUs. In South Korea, the distribution networks are relatively more complicated than the transmission networks. Furthermore, there are many distribution facilities compared to the number of facilities in the transmission networks. Thus, the wireless hopping method is applied between DCTUs to reduce the costs for establishing the communication network.

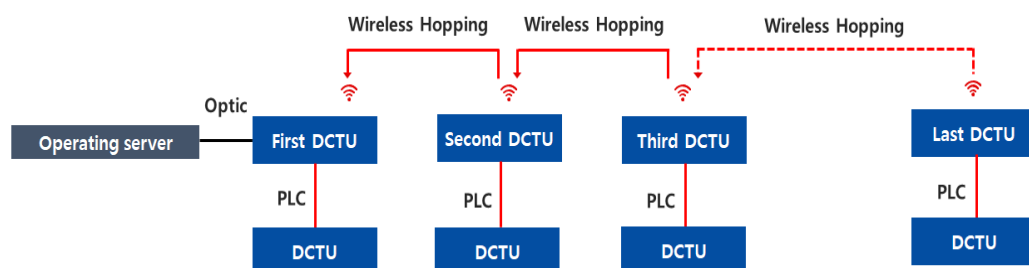


Figure 5. Schematic of wireless hopping and optic communication.

The optic fiber communication is used between the operating server and the first DCTU. The first DCTU closest to the operating server transmits the collected data to the operating server using optic fiber. The reason for using optic transmission instead of the wireless hopping method between the first DCTU and operating server is to send the collected data to the operating server promptly.

The operating server performs the PD online detection algorithm to identify PD faults continuously in real-time. Algorithm 1 describes the PD online detection algorithm. First, the proposed model decides whether the data are PD or noise when the newly transmitted data are acquired from the online distribution network. When the data are PD, the PD magnitude is recorded, and the PD count is increased. Then, the cumulative PD magnitude and counts of the distribution facility are checked. When the cumulative magnitude and counts surpass the setting value, the PD detection system alerts the alarm and requests facility maintenance. However, when the data are noise, the noise data are discarded.

Algorithm 1 PD online detection algorithm.

```

1: m_limit ← The setting value of cumulative PD_magnitudes
2: c_limit ← The setting value of cumulative PD_counts

3: function PROPOSED_MODEL()
4:   if new data is available then
5:     Proposed model decides whether the data is PD or Noise
6:   end if
7: end function

8: while output of the proposed_model is available do
9:   if output is PD then
10:    Record the PD_magnitudes
11:    PD_counts ← PD_counts + 1
12:    Check the cumulative PD_magnitude and PD_counts of the facility
13:    if (cumulative_magnitudes ≥ m_limit) and (cumulative_counts ≥ c_limit) then
14:      Alert the alarm
15:      Request the facility maintenance
16:    end if
17:   else
18:     Noise data is discarded
19:   end if
20: end while

```

3.2. PD and Noise Data Patterns

Phase-resolved partial discharge (PRPD) pattern is used as the input data of deep learning models for PD detection. The PRPD pattern is defined using three parameters: apparent charge (Q), phase angle (ϕ), and the rate of occurrence (n) in a specific time based on IEC-60270 standard [1]. An example of the PRPD pattern acquired from the PD online detection system is shown in Figure 6. The red clustering PD patterns occur at the rise and falling cycles within 1/60 s. The red clustering patterns indicate many PD occurrences within a specific time. However, the noise pattern differs from the PRPD pattern, as shown in Figure 7. The noise pattern does not show red clustering patterns within a cycle. Such PRPD patterns have been collected from PD online detection systems continuously in real-time.

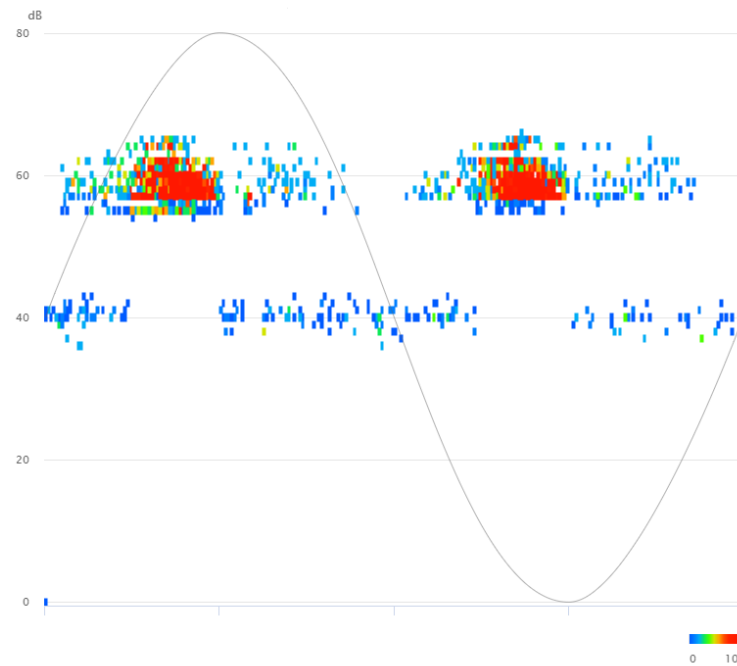


Figure 6. PRPD pattern sample acquired from PD online detection system.

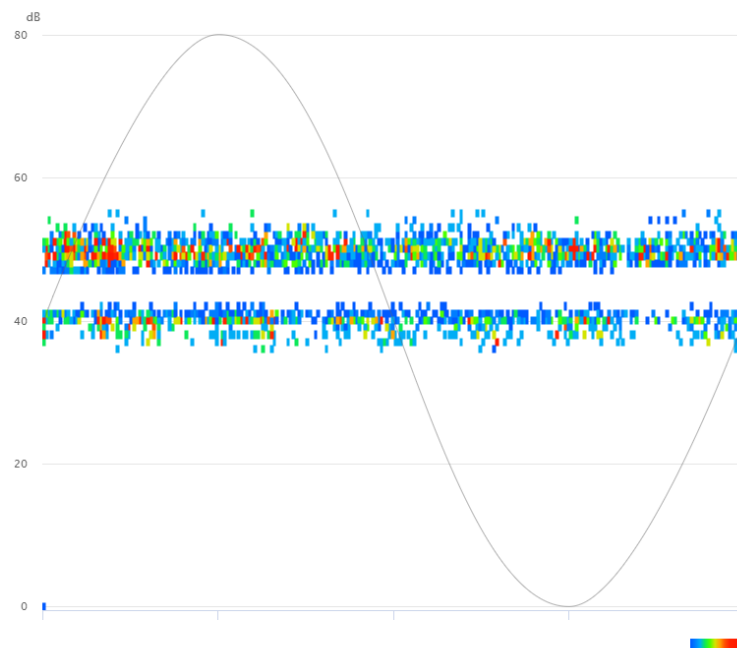


Figure 7. Noise pattern sample acquired from PD online detection system.

3.3. Proposed Model

In this study, a CNN transfer-learning model is proposed for PD detection. In [24], the transfer learning is defined as a domain adaptation model in the situation where it has learned in one setting to improve generalization in another setting. There are several reasons for using the transfer-learning model. First, the PD studies in [19,25–27] reported that the transfer-learning models can improve accuracy using few data compared with other benchmark models such as traditional machine learning models.

Furthermore, in [28], the transfer-learning model can allow developers to circumvent the need for many new data. A model that has been trained on a task with many labeled

training data can handle a new but similar task with far fewer data. Furthermore, using a pretrained model often accelerates the process of training the model on a new task, resulting in a more accurate and effective model.

In the current case, few real data obtained from the distribution network could not exhibit the high possibility of improving accuracy for the deep learning model. It is not easy to obtain enough real data in the distribution network. Thus, the transfer-learning model can be an effective method for improving accuracy efficiently using few data. The step of designing transfer-learning is described as follows:

- Introducing a previously trained transfer-learning model.
- Freezing them to avoid destroying any of the information for the model.
- Adding some new training layers on top of the frozen layers.
- Training the new layers using a new dataset.
- Fine-tuning, which is the process of adjusting parameters, can achieve meaningful improvements by adjusting parameters.

The first proposed transfer-learning model is a ResNet deep learning model [27]. The ResNet model has the advantage of designing a deeper network model than the ordinary CNN deep learning model. When the deep learning model becomes a deeper network, a degradation problem is exposed. When the network depth increases, the accuracy gets saturated; i.e., adding more layers to a deep network can lead to higher training error.

The reason for the higher training error is described as follows. When the deep neural networks update as a method of backpropagation proportional to the partial derivative of the error function in each training iteration, the gradient could be vanishingly small. Thus, the longer gradient chain can have a high possibility of causing the gradient-vanishing problem because the gradient could become small whenever the chain becomes deeper.

Furthermore, backpropagation computes gradients using the chain rule. The concept of backpropagation using chain rule is described in Figure 8. When the layers become deeper, the partial derivative chain for network updates can become longer. The longer chain can increase the computational complexity of the training model. The computational complexity can also cause higher training errors.

To solve these problems, a deep residual learning (ResNet) model is introduced. The ordinary layer architecture of the deep learning model is described in Figure 9. The ResNet model has different layer architecture compared with the ordinary architecture. The ResNet architecture can be realized with shortcut connections. The layer architecture with the shortcut connection is shown in Figure 10. The derivative of the formulation ($f(x) + x$) is always larger than the output value ($=1$). Although the deep learning model becomes a deeper network, the gradient could not vanish since the gradient is always larger than one ($=1$). Thus, the gradient-vanishing problem can be solved using the ResNet framework.

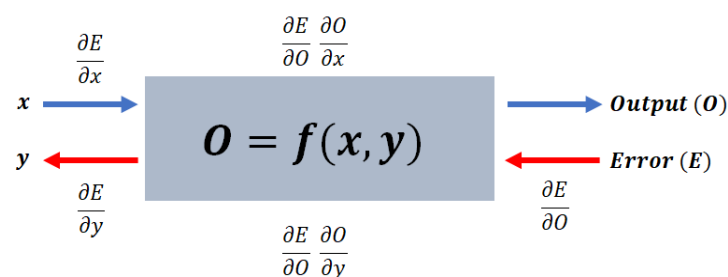


Figure 8. Concept of backpropagation using chain rule.

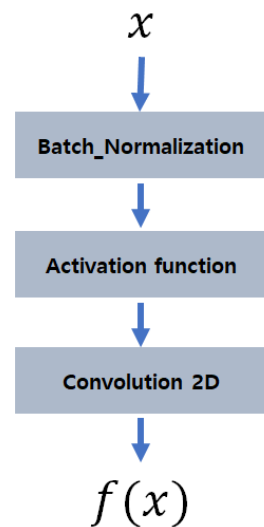


Figure 9. Ordinary layer architecture of deep learning model.

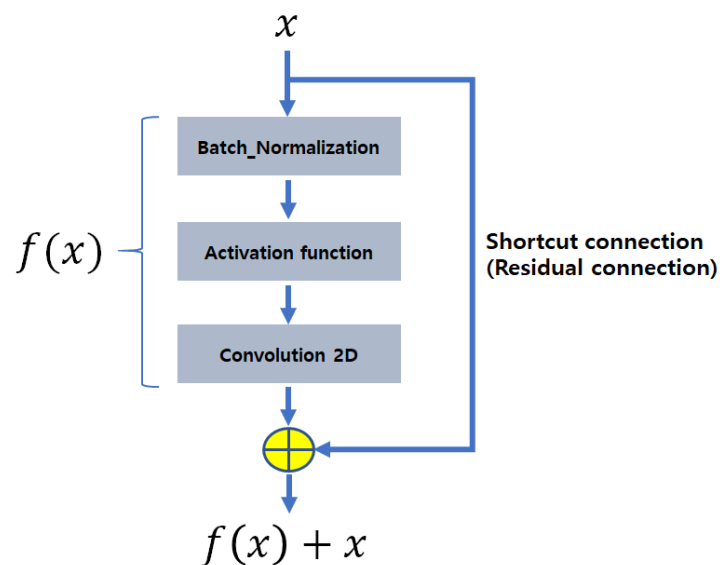


Figure 10. Layer architecture of ResNet model with shortcut connection.

In this study, the ResNet50V2 model is modified. The modified transfer-learning models for identifying best performance model are classified into three models as follows:

- ResNet50V2_C1: Baseline frozen model;
- ResNet50V2_C2: The fine-tuning model with modified layers closest to input;
- ResNet50V2_C3: The fine-tuning model with modified layers closest to output.

Figure 11 shows three ResNet50v2 models. The baseline frozen model (ResNet50V2_C1) is a transfer-learning model used to freeze all layers or parameters, except dense layers at the end of layers. The second proposed model (ResNet50V2_C2) is a fine-tuning model with modified layers closest to input, including dense layers. The third proposed model (ResNet50V2_C3) is a fine-tuning model with modified layers closest to the output, including dense layers.

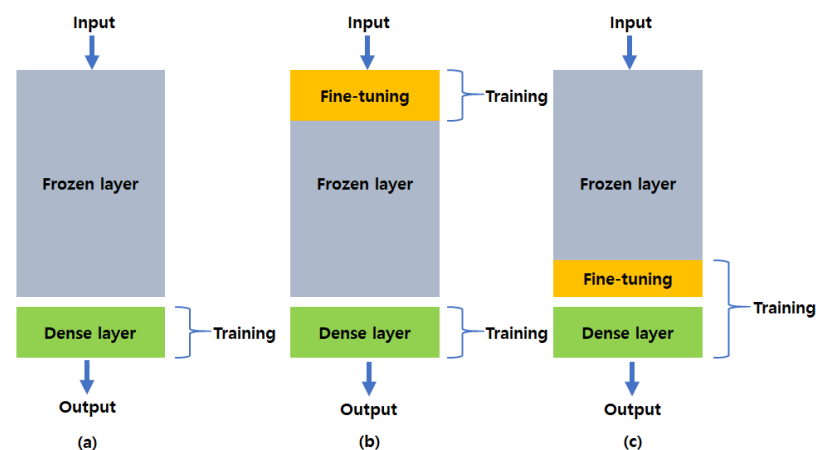


Figure 11. Three kinds of modified transfer-learning models: (a) ResNet50V2_C1 and MobileNet50V2_C1, (b) ResNet50V2_C2 and MobileNet50V2_C2, (c) ResNet50V2_C3 and MobileNet50V2_C3.

Regarding the dataset split, we use ImageDataGenerator library to divide the training dataset and test dataset. The software library has been provided by tensorflow keras framework. Using the library, we can easily split the dataset into the training and the test datasets in various proportions fairly and randomly.

In addition, the reason to use fewer training data than the test data is that we try to make the proposed model become more appropriate model using small real PD data only based on the fact that the real PD dataset is still much smaller than the dataset commonly used in computer vision studies as shown in [26]. Considering this, we apply validation split (0.8) parameter to the proposed model using ImageDataGenerator library.

The validation split (0.8) means the training data (20%) and the test data (80%) in the dataset. We use the validation parameter (0.8) to verify the effectiveness of the proposed model developed by small real PD test dataset only using much more test dataset. Thus, we try to prove the robustness of the proposed model in case the model is developed by using small real PD data.

The main parameters for designing the proposed model are justified based on [24]. Regarding the main parameters in Table 2, the epochs mean the number of training iterations over the dataset. The input shape parameter means the input size, and we use three-dimensional data as a model input. The batch size means the total number of training samples present in a single batch.

Three ResNet models are designed using main parameters, as shown in Table 2. The main parameters of ResNet models are 126 training samples, 497 test samples, 5 epochs, (224,224,3) input shape, and 16 batch sizes.

Three ResNet models have different trainable layers. The ResNet50V2_C1 has trainable dense layers only closest to the output. The ResNet50V2_C2 has fine-tuning layers from 0 to 15 layers and trainable dense layers. The ResNet50V2_C3 has fine-tuning layers from 176 to last layers and trainable dense layers.

Table 2. Main parameters for designing proposed ResNet models.

Parameter	Description
Training samples	126
Test samples	497
Epochs	5
Input_shape	(224,224,3)
Batch_size	16

The second proposed transfer-learning model is MobileNet deep learning model [29]. The MobileNet model has the advantage of reducing the computational costs compared to standard CNN models. The MobileNet model employs two types of convolution: depthwise and pointwise convolutions.

First, the description of the standard convolution for calculating the computational costs is shown in Figure 12. The description of the depthwise and pointwise convolutions for computational costs is shown in Figure 13.

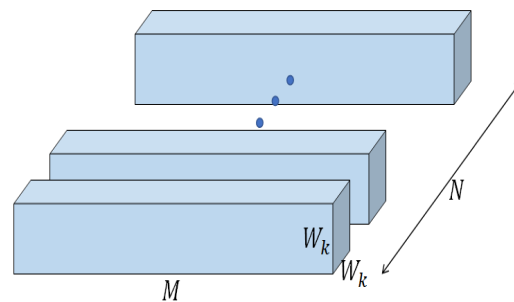


Figure 12. Standard convolution for deep learning model.

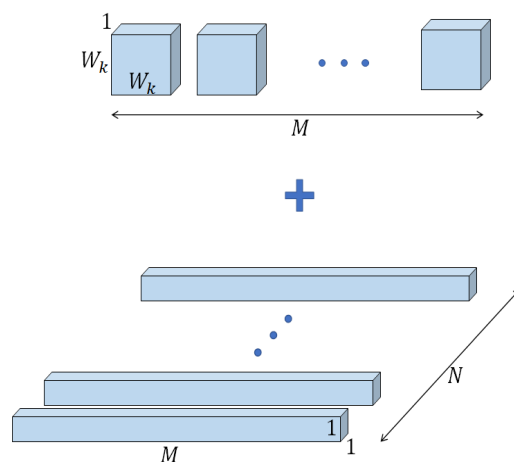


Figure 13. Depthwise and pointwise convolutions for MobileNet model.

As a result of the computational costs, the MobileNet can reduce computation complexity using depthwise and pointwise convolution. Standard convolution has the computational costs as follows:

$$W_k \times W_k \times W_f \times W_f \times M \times N \quad (1)$$

where W_k is the width and height of kernel filter, W_f is the width and height of feature map, M is the size of input channel, and N is the size of output channel.

Depthwise and pointwise convolution of MobileNet has the computational costs is as follows:

$$W_k \times W_k \times W_f \times W_f \times M + W_f \times W_f \times M \times N \quad (2)$$

$$\frac{W_k \times W_k \times W_f \times W_f \times M + W_f \times W_f \times M \times N}{W_k \times W_k \times W_f \times W_f \times M \times N} = \frac{1}{N} + \frac{1}{W_k^2} \quad (3)$$

Equation (3) explains the ratio of the computational costs between the standard convolution and depthwise and pointwise convolution. N is much bigger than D_k . When the D_k is three, the computation can be reduced to 1/9.

Three modified MobileNetV2 models are designed using parameters, as shown in Table 3. As shown in Figure 11, the MobileNetV2 model is classified into three models as follows:

- MobileNetV2_C1: Baseline frozen model
- MobileNetV2_C2: The fine-tuning model with modified layers closest to input
- MobileNetV2_C3: The fine-tuning model with modified layers closest to output

The main parameters are 126 training samples, 497 test samples, 5 epochs, (224,224,3) input shape, and 16 batch sizes. Moreover, three MobileNet models have different trainable layers. The MobileNetV2_C1 has trainable dense layers only closest to the output. The MobileNet50V2_C2 has fine-tuning layers from 0 to 4 layers and trainable dense layers. The ResNet50V2_C3 has fine-tuning layers from 45 to last layers and trainable dense layers.

Table 3. Main parameters for designing proposed MobileNet models.

Parameter	Description
Training samples	126
Test samples	497
Epochs	5
Input_shape	(224,224,3)
Batch_size	16

3.4. Benchmark Model

In this study, there are two benchmark models: CNN and SVM models. The first benchmark model (CNN) has the convolutional operation and pooling technique for extracting feature data. The convolutional operation is used as a signal processing method. Recently, the convolutional operation has been used in the computer vision domain. The convolutional operation is described as follows:

$$\text{Convolutional_operation}(i, j) = \sum_x \sum_y \text{Image}(x, y) \text{Kernel}(i - x, j - y) \quad (4)$$

The advantages of convolutional operation are as follows:

- Sparse interaction,
- Fewer parameter operation, and
- Lower computation with efficiency.

When a kernel filter moves with stride within an input image, the element-wise convolutional operation is performed. The operation contributes to sparse interaction and fewer parameter operations. Moreover, the operation can be used to extract meaningful features. The sparse and fewer calculations can save memory and lower computation complexity. Furthermore, a kernel filter can share parameters for finite computation.

The pooling can be defined as summary statistics of the nearby outputs at a certain location between an input image and kernel filter. The pooling is divided into two types: max and average pooling. Max pooling means a maximum value within a kernel filter on an input image. Average pooling means an average value within a kernel filter. The advantage of max or average pooling is to show the robustness of invariant output against small input change.

The main parameters for designing the benchmark CNN models are also justified based on [24]. In particular, in Table 4, the dropout means the sub-networks that can be formed by removing non-output units from a base network. The activation is a function to design the output node used in artificial neural networks. As a kind of activation function, relu is a popular linear function that will output the input directly if it is positive; otherwise, it will output zero. The softmax function is the last activation function to normalize the output of a network. The cross-entropy is commonly used in machine learning as a loss function. The adam is also a popular optimization algorithm using adaptive learning rate.

A description of the CNN benchmark models (CNN_2 layers and CNN_4 layers) is provided in Table 4. The main parameters of the CNN models are 2 (or 4) layers, 126 training samples, 497 test samples, 5 epochs, (224,224,3) input shape, (2,2) max pooling, 0.25 dropout, relu activation for convolution layer, softmax for dense layer, categorical crossentropy loss function, and adam optimizer.

Table 4. Main parameters for designing benchmark CNN models.

Parameter	Description
Layers	2 (or 4)
Training samples	126
Test samples	497
Epochs	5
Input_shape	(224,224,3)
Max_pooling	(2,2)
Dropout	0.25
Activation function	relu (convolution layer) softmax (dense layer)
Loss function	categorical_crossentropy
Optimizer	adam

The second benchmark model is an SVM model. As shown in Figure 14, the SVM model introduces the margin concept. The margin is the distance between the decision surface and the closest data point. When the margin is bigger, the model could be better. Thus, the optimization function of the SVM model is to maximize margin. The optimization function maximizes the margin subject for all training data to be located or behind the margin. The optimized function is formulated as follows:

$$\text{minimize: } \frac{1}{2} \|\omega\|^2 \quad (5)$$

$$\text{subject to: } y_i(w^T x_i + b) - 1 \geq 0 \text{ for all } i \quad (6)$$

Soft margin SVM can allow the amount of error for each training data. The modified objective function is given as follows:

$$\text{minimize: } \frac{1}{2} \|\omega\|^2 + C \sum_i \xi_i \quad (7)$$

$$\text{subject to: } y_i(w^T x_i + b) - 1 + \xi_i \geq 0, \xi_i \geq 0 \text{ for all } i \quad (8)$$

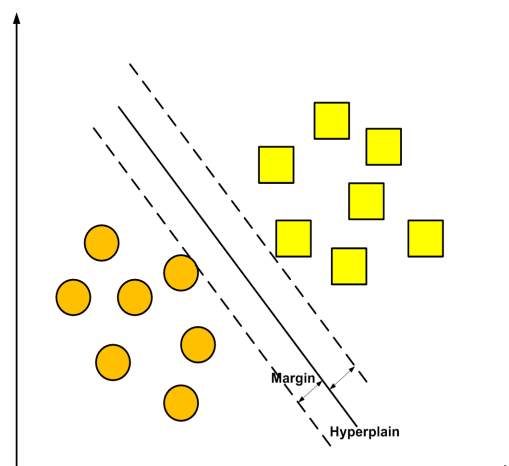


Figure 14. The graphical concept for SVM model.

Nonlinear SVM can allow high-dimensional mapping for each training data. The nonlinear SVM moves the data into another feature space. It is theoretically formulated as follows:

$$x = (x_1, \dots, x_n) \rightarrow \varphi(x) = (\varphi_1(x), \dots, \varphi_n) \quad (9)$$

$$\text{minimize: } \frac{1}{2} \|\omega\|^2 + C \sum_i \xi_i \quad (10)$$

$$\text{subject to: } y_i(w^T \varphi(x_i) + b) - 1 + \xi_i \geq 0, \xi_i \geq 0 \text{ for all } i \quad (11)$$

$$x = (x_1, \dots, x_n) \rightarrow \varphi(x) = (\varphi_1(x), \dots, \varphi_n) \quad (12)$$

$$\text{minimize: } \frac{1}{2} \|\omega\|^2 + C \sum_i \xi_i \quad (13)$$

$$\text{subject to: } y_i(w^T \varphi(x_i) + b) - 1 + \xi_i \geq 0, \xi_i \geq 0 \text{ for all } i \quad (14)$$

The main parameters for designing the SVM can be justified based on [24,30]. The radial basis function (RBF) kernel is used as a nonlinear mapping function of the SVM model. The RBF kernel is defined as follows:

$$\text{Rbf_kernel_function} = \exp(-\gamma \|x_1 - x_2\|^2) \quad (15)$$

As shown in Equations (7) and (10), C denotes a tradeoff parameter for controlling errors. As shown in Equation (15), γ is a parameter for deciding the boundary as the inverse of the radius of RBF kernel.

When two parameters (C and γ) are large, the model becomes more complicated. The large parameters (C and γ) could lead to an overfitting model. Thus, it is essential to apply appropriate values of C and γ to the SVM model.

Table 5 presents the main parameters of the SVM benchmark model. The main parameters of SVM are 126 training samples and 313 test samples. We use default parameters C (=1) and γ (=auto) for RBF kernel function.

Table 5. Main parameters for designing an SVM benchmark model.

Parameter	Description
Training samples	126
Test samples	313
Kernel	rbf
C	1
γ	auto

4. PD Detection Accuracy Results

In this section, two kinds of proposed models (ResNet50V2 and MobileNet50V2) are analyzed to determine the effects and advantages of the transfer-learning model to develop the deep learning model using few data (126 training samples) for PD detection. Moreover, three benchmark models are examined to verify the PD detection performance compared with the proposed models.

The proposed model is a binary classification model to decide whether the data are PD or Noise. A confusion matrix is usually used as performance metrics. Performance metrics can be presented as follows:

- True Positive (TP): The data are also PD when the proposed model judges as a PD.
- True Negative (TN): The data are also Noise when the proposed model judges as a Noise.
- False Positive (FP): The data are Noise when the proposed model judges as a PD.

- False Negative (FN): The data are PD when the proposed model judges as a Noise. The accuracy for confusion matrix can be calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (16)$$

Table 6 presents the accuracy results of proposed transfer-learning and benchmark models for PD detection. As shown in Table 6, ResNet50V2_C1 and MobileNet50V2_C1 provide the highest training accuracy (100%) among the six proposed models.

ResNet50V2_C3 has the highest accuracy (100% training accuracy and 96.2% test accuracy) among ResNet50V2 models, while the MobileNet50V2_C3 model has the best accuracy (99.2% training accuracy and 97.4% test accuracy) among MobileNet50V2 models.

ResNet50V2_C3 and MobileNet50V2_C3 achieve the highest test score (96.2% and 97.4%, respectively) compared with benchmark models (CNN and SCM).

The benchmark models (CNN_2 layer and CNN_4 layer) show as results 100% training and 89% test scores. The SVM model achieves the lowest accuracy results (50% training and 49.8% test scores).

These results showed that the proposed transfer-learning models (ResNet50V2_C3 and MobileNet50V2_C3) improve the accuracy results compared with other benchmark models using few data. The MobileNet50V2_C3 model can provide the highest test accuracy (97.4%) compared with any other models using previously unseen test data.

Moreover, the transfer-learning model can be an effective PD detection model for on-site low voltage distribution network since the proposed models can be developed using few real PD data (126 training samples), and the proposed transfer-learning models show superior accuracy results than other benchmark models.

Table 6. Accuracy results of proposed transfer-learning models and benchmark models for PD detection.

Model	Training Accuracy	Test Accuracy
ResNet50V2_C1	100%	89.1%
ResNet50V2_C2	73.0%	51.5%
ResNet50V2_C3	100%	96.2%
MobileNet50V2_C1	100%	88.7%
MobileNet50V2_C2	51.5%	16.7%
MobileNet50V2_C3	99.2%	97.4%
CNN_2 layers	100%	89.6%
CNN_4 layers	100%	89.3%
SVM	50.0%	49.8%

Figures 15 and 16 show the training accuracy results of ResNet50V2 and MobileNet50V2 each epoch. First, C1 (baseline frozen model) and C3 (fine-tuning model with modified layers closest to output) models are more effective for improving accuracy than the C2 model (fine-tuning model with modified layers closest to input).

As shown in Figures 15 and 16, when C1 and C3 models of ResNet50V2 and MobileNet50V2 reach epoch 2, the accuracy almost becomes 100%. These results indicate that the models become overfit models when the epoch counts increase. Thus, the model's epoch, which is the number of the learning process the deep learning algorithm has completed is limited to five (=5). In the proposed system, the optimized model architecture and weights are stored after each epoch to obtain the best model for PD detection. Then, the best model architecture and weights are extracted from the whole storing results after each epoch.

As shown in Figures 17 and 18, the test accuracy results of the proposed models can be identified. The C1 and C3 models also provide better accuracy than the C2 model. The C3 model shows the highest accuracy among the models. This result indicates that the

fine-tuning closest to output play a significant role in improving the accuracy performance of the deep learning model.

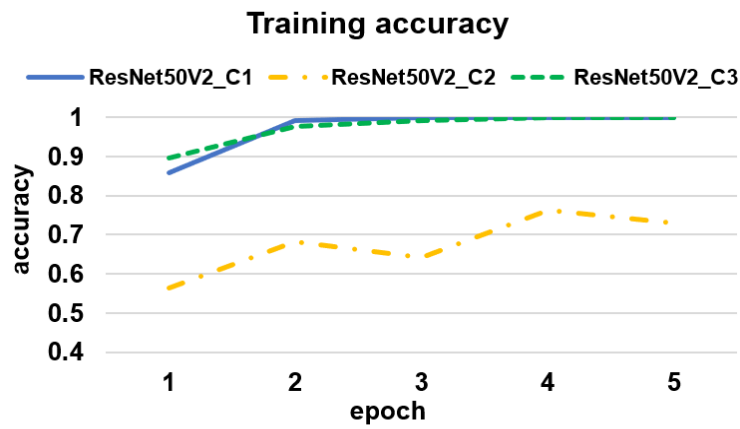


Figure 15. Training accuracy results of proposed ResNet50V2 models.

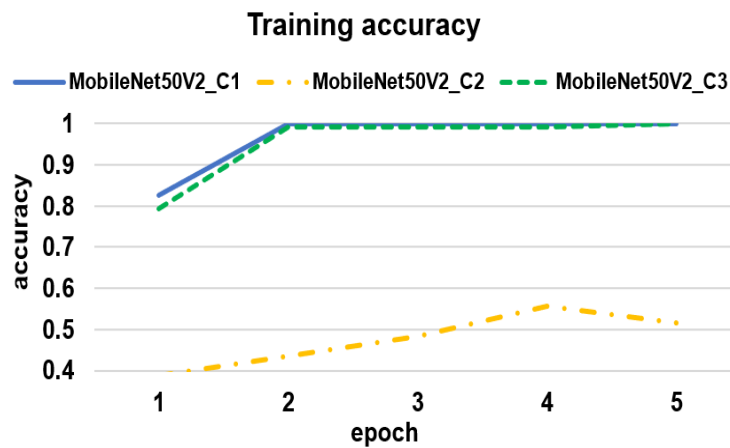


Figure 16. Training accuracy results of proposed MobileNet50V2 models.

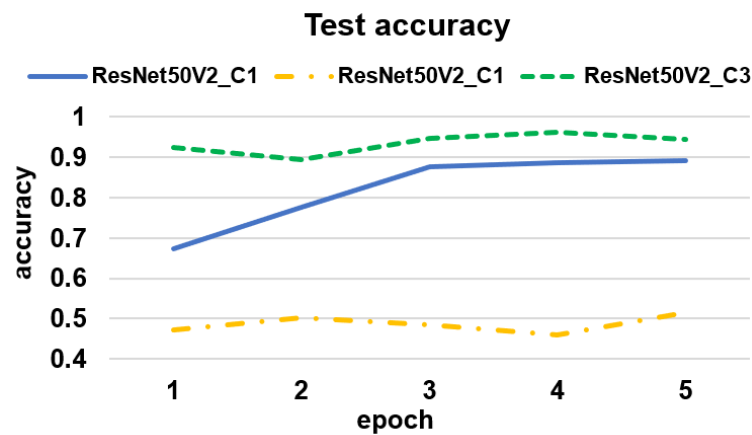


Figure 17. Test accuracy results of proposed ResNet50V2 models.

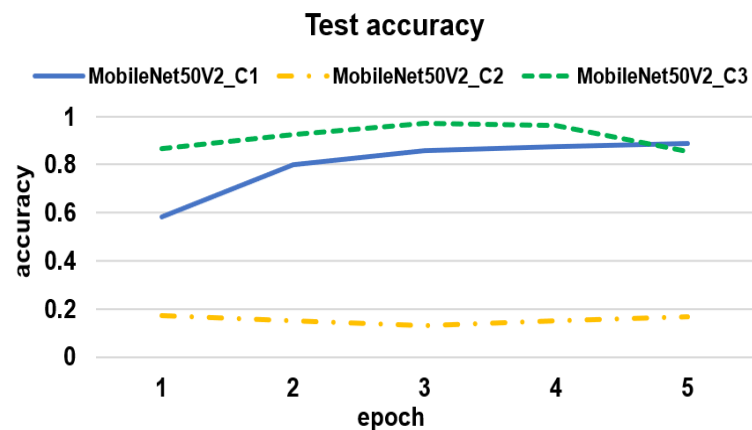


Figure 18. Test accuracy results of proposed MobileNet50V2 models.

5. Discussion

The effectiveness of the proposed PD detection system is summarized as follows:

- The PD online detection system was proposed for long-term operational sustainability of on-site low voltage distribution network since there are few studies of PD fault detection for a low voltage distribution network.
- The automated PD online detection system can obtain real PD data through the continuous monitoring of PD occurrence.
- The effectiveness of the proposed transfer-learning models was verified based on the fact that the CNN transfer-learning models developed using small real PD data (126 training samples only) showed improved test accuracy for real PD data (497 test samples) compared with other benchmark models.
- The proposed transfer-learning models (ResNet50V2_C3 and MobileNet50V2_C3) achieved the highest test score (96.2% and 97.4%, respectively) compared with benchmark models (CNN and SVM).

6. Conclusions

In this study, PD online detection system and models for long-term operational sustainability of on-site low voltage distribution network were proposed using CNN transfer-learning. At first, the proposed system makes it possible to acquire as many real PD data as possible through the continuous monitoring of PD occurrence using the PD online detection system. Secondly, the modified transfer-learning models were proposed as the PD detection model. The PD detection accuracy results of the proposed models showed that the proposed CNN transfer-learning models (ResNet50V2_C3 and MobileNet50V2_C3) are more effective models to show the improved accuracy (96.2 % and 97.4%, respectively) than benchmark models such as CNN and SVM. The models developed using few real PD data (126 training samples only) showed improved test accuracy for real PD data (497 test samples) compared with other benchmark models.

In future study, first, the PD online detection system will obtain continuously more real PD data to classify the acquired real PD data into various kinds of PDs (void, surface, and corona) for long-term sustainability of on-site low voltage distribution networks. Second, the proposed transfer-learning models for PD detection will be developed to improve the accuracy performance and diagnose various PDs using only real PD data.

Author Contributions: J.K. designed the proposed system and model, performed various experiments, analyzed the results, and wrote the manuscript. K.-I.K. supervised this research and reviewed the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education (2018R1D1A1B07043731) and Institute for Information & Communications Technology Planning & Evaluation (IITP) grant funded

by the Korea government (MSIT) (No. 2019-0-01343, Training Key Talents in Industrial Convergence Security).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available on request.

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

References

- Duan, L.; Hu, J.; Zhao, G.; Chen, K.; He, J.; Wang, S.X. Identification of Partial Discharge Defects Based on Deep Learning Method. *IEEE Trans. Power Deliv.* **2019**, *34*, 1557–1568. [[CrossRef](#)]
- Peng, X.; Yang, F.; Wang, G.; Wu, Y.; Li, L.; Li, Z.; Bhatti, A.A.; Zhou, C.; Hepburn, D.M.; Reid, A.J.; et al. A Convolutional Neural Network-Based Deep Learning Methodology for Recognition of Partial Discharge Patterns from High-Voltage Cables. *IEEE Trans. Power Deliv.* **2019**, *34*, 1460–1469. [[CrossRef](#)]
- Dai, J.; Teng, Y.; Zhang, Z.; Yu, Z.; Sheng, G.; Jiang, X. Partial Discharge Data Matching Method for GIS Case-Based Reasoning. *Energies* **2019**, *12*, 3677. [[CrossRef](#)]
- Wang, Y.; Yan, J.; Yang, Z.; Liu, T.; Zhao, Y.; Li, J. Partial Discharge Pattern Recognition of Gas-Insulated Switchgear via a Light-Scale Convolutional Neural Network. *Energies* **2019**, *12*, 4674. [[CrossRef](#)]
- Kweon, D.J.; Chin, S.B.; Kwak, H.R.; Kim, J.C.; Song, K.B. The Analysis of Ultrasonic Signals by Partial Discharge and Noise From the Transformer. *IEEE Trans. Power Deliv.* **2005**, *20*, 1976–1983. [[CrossRef](#)]
- Hikita, M.; Ohtsuka, S.; Wada, J.; Okabe, S.; Hoshino, T.; Maruyama, S. Study of partial discharge radiated electromagnetic wave propagation characteristics in an actual 154 kV model GIS. *IEEE Trans. Dielectr. Electr. Insul.* **2012**, *19*, 8–17. [[CrossRef](#)]
- Samuelsson, O.; Hemmingsson, M.; Nielsen, A.; Pedersen, K.; Rasmussen, J. Monitoring of Power System Events at Transmission and Distribution Level. *IEEE Trans. Power Syst.* **2006**, *21*, 1007–1008. [[CrossRef](#)]
- Barrios, S.; Buldain, D.; Comech, M.; Gilbert, I.; Orue, I. Partial Discharge Classification Using Deep Learning Methods—Survey of Recent Progress. *Energies* **2019**, *12*, 2485. [[CrossRef](#)]
- Iorkyase, E.T.; Tachtatzis, C.; Glover, I.A.; Lazaridis, P.; Upton, D.; Saeed, B.; Atkinson, R.C. Improving RF-Based Partial Discharge Localization via Machine Learning Ensemble Method. *IEEE Trans. Power Deliv.* **2019**, *34*, 1478–1489. [[CrossRef](#)]
- Zhang, X.; Shi, M.; Cai, J.; Li, J. A Novel Partial Discharge Detection Method for Power Transformers On Site Adopting Its Component as Ultra-High Frequency Sensor. *IEEE Trans. Power Deliv.* **2019**, *34*, 2269–2271. [[CrossRef](#)]
- Kong, W.; Men, K.; Guo, Z.; Dong, Z.Y.; Zhang, R.; Jia, Y. Automatic Online Partial Discharge Diagnosis via Deep Learning. In Proceedings of the 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 17–20 February 2020; pp. 1–5. [[CrossRef](#)]
- Balouji, E.; Hammarstrom, T.; McKelvey, T. Partial Discharge Classification in Power Electronics Applications using Machine Learning. In Proceedings of the 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Ottawa, ON, Canada, 11–14 November 2019; pp. 1–5. [[CrossRef](#)]
- Yeo, J.; Jin, H.; Yuen, C.; Ng, C.S. Developing Neural Network Models for Partial Discharge Analysis. In Proceedings of the IECON 2020 46th Annual Conference of the IEEE Industrial Electronics Society, Singapore, 18–21 October 2020; pp. 1573–1577. [[CrossRef](#)]
- Wan, X.; Song, H.; Luo, L.; Li, Z.; Sheng, G.; Jiang, X. Pattern Recognition of Partial Discharge Image Based on One-dimensional Convolutional Neural Network. In Proceedings of the 2018 Condition Monitoring and Diagnosis (CMD), Perth, Australia, 23–26 September 2018; pp. 1–4. [[CrossRef](#)]
- Banno, K.; Nakamura, Y.; Fujii, Y.; Takano, T. Partial Discharge Source Classification for Switchgears with Transient Earth Voltage Sensor Using Convolutional Neural Network. In Proceedings of the 2018 Condition Monitoring and Diagnosis (CMD), Perth, Australia, 23–26 September 2018; pp. 1–5. [[CrossRef](#)]
- Puspitasari, N.; Khayam, U.; Suwarno; Kakimoto, Y.; Yoshikawa, H.; Kozako, M.; Hikita, M. Partial Discharge Waveform Identification using Image with Convolutional Neural Network. In Proceedings of the 2019 54th International Universities Power Engineering Conference (UPEC), Bucharest, Romania, 3–6 September 2019; pp. 1–4. [[CrossRef](#)]
- Zhang, A.; He, J.; Lin, Y.; Li, Q.; Yang, W.; Qu, G. Recognition of partial discharge of cable accessories based on convolutional neural network with small data set. *COMPEL Int. J. Comput. Math. Electr. Electron. Eng.* **2020**, *39*, 431–446. [[CrossRef](#)]
- Tuyet-Doan, V.N.; Do, T.D.; Tran-Thi, N.D.; Youn, Y.W.; Kim, Y.H. One-Shot Learning for Partial Discharge Diagnosis Using Ultra-High-Frequency Sensor in Gas-Insulated Switchgear. *Sensors* **2020**, *20*, 5562. [[CrossRef](#)]
- 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](#)]
- Liu, N.; Fan, B.; Xiao, X.; Yang, X. Cable Incipient Fault Identification with a Sparse Autoencoder and a Deep Belief Network. *Energies* **2019**, *12*, 3424. [[CrossRef](#)]
- Zhou, X.; Wu, X.; Ding, P.; Li, X.; He, N.; Zhang, G.; Zhang, X. Research on Transformer Partial Discharge UHF Pattern Recognition Based on Cnn-Istm. *Energies* **2019**, *13*, 61. [[CrossRef](#)]

22. Florkowski, M. Classification of Partial Discharge Images Using Deep Convolutional Neural Networks. *Energies* **2020**, *13*, 5496. [[CrossRef](#)]
23. 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](#)]
24. Goodfellow, I.J.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
25. Liu, T.; Yan, J.; Wang, Y.; Du, Y. Research on GIS Partial Discharge Pattern Recognition Based on Deep Residual Network and Transfer Learning in Ubiquitous Power Internet of Things Context. In Proceedings of the 2020 5th Asia Conference on Power and Electrical Engineering (ACPEE), Chengdu, China, 4–7 June 2020; pp. 207–211. [[CrossRef](#)]
26. Tang, Z.; Cao, Z. Application of Convolutional Neural Network Transfer Learning in Partial Discharge Pattern Recognition. In Proceedings of the 2020 IEEE International Conference on High Voltage Engineering and Application (ICHVE), Beijing, China, 6–10 September 2020; pp. 1–4. [[CrossRef](#)]
27. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity Mappings in Deep Residual Networks. *arXiv* **2016**, arXiv:1603.05027.
28. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *arXiv* **2019**. arXiv:1911.02685.
29. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; Adam, H. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv* **2017**, arXiv:1704.04861.
30. Hao, L.; Lewin, P.L.; Swingler, S.G. Improving detection sensitivity for partial discharge monitoring of high voltage equipment. *Meas. Sci. Technol.* **2008**, *19*, 055707. [[CrossRef](#)]