Received 19 December 2023, accepted 1 February 2024, date of publication 12 February 2024, date of current version 19 April 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3365135

RESEARCH ARTICLE

Classification of PRPD Pattern in Cast-Resin Transformers Using CNN and Implementation of Explainable AI (XAI) With Grad-CAM

HO-SEUNG KIM [1], JIHO JUNG [1], RYUL HWANG [1], SEONG-CHAN PARK [2], SEUNG-JAE LEE [2], GYU-TAE KIM [2], AND BANG-WOOK LEE [1], (Senior Member, IEEE)

[1] Department of Electrical Engineering, Hanyang University Erica, Ansan 15588, South Korea
[2] DS Division, Samsung Electronics Company Ltd., Hwaseong 18448, South Korea

Corresponding author: Bang-Wook Lee (bangwook@hanyang.ac.kr)

This work was supported in part by the DS Division, Samsung Electronics, under Grant IO221118-03611-01; and in part by Korea Institute of Energy Technology Evaluation and Planning (KETEP) Grant funded by Korea Government (MOTIE) (DC Grid Energy Innovation Research Center) under Grant 20224000000160.

ABSTRACT Cast-resin transformers are affected by deterioration due to manufacturing defects and continuous load. Studying PD, which is capable of detecting defects or degradation in advance, is important. With the rapid advancement of AI technologies, research on PD classification using CNN models is being actively conducted. However, due to the black box problem, it is impossible to explain the reasoning behind the learning outcomes. Therefore, relying solely on predictive outcomes of learning for PD classification raises issues of reliability. Recent studies in various fields are progressing with the application of XAI to address the black box issue of CNNs, aiming to identify the criteria used for making predictions. However, research on applying XAI in AI-based PD classification is currently insufficient. Therefore, further study on the implementation of XAI is necessary. In this paper, an excellent CNN model was applied to image classification for PD classification of cast-resin transformers, and the grad-cam model was used for XAI. This approach proposes a method for humans to comprehend the rationale behind the learning outcomes. The training data includes artificial defects created in laboratory settings and noise captured in cast-resin transformers using UHF sensors. Our research demonstrated that PD and noise due to defects can be identified with an accuracy of approximately 97%. The reasons for successful and failed results were analyzed through XAI. Consequently, it was observed that the application of XAI to CNN models leads to the construction of a more reliable model.

INDEX TERMS Cast-resin transformer, PD, pattern classification, convolution neural network (CNN), explainable artificial intelligence (XAI), gradient weighted class activation mapping (Grad-CAM).

I. INTRODUCTION

Cast-resin transformers, in contrast to traditional oil-filled transformers, use epoxy resin to cast their core and windings, eliminating the need for insulating oil. This leads to reduced maintenance requirements due to the absence of oil replacement and offers the advantage of lower fire risk, making them suitable for indoor installations [1], [2]. Additionally, they are less susceptible to moisture ingress and have a high resistance to pollutants, making them environmentally safer [3]. Therefore, the demand for cast-resin transformers is increasing in line with the continual growth in power consumption [4].

However, cast-resin transformers have a risk of accidents leading to insulation breakdown due to manufacturing defects such as cracks, heat accumulation, and voids [5], [6]. Initial defects can significantly shorten the lifetime of cast-resin

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

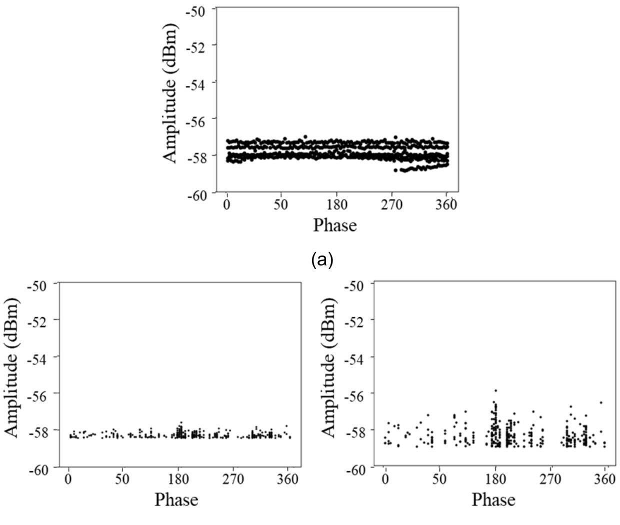
transformer and greatly affect its reliability [7]. Recently, while the incidence of accidents due to manufacturing defects has decreased, there is an increase in degradation due to thermal, mechanical, and electrical caused by continuous loads [8], [9]. Therefore, research on PD for early detection and analysis of accidents caused by manufacturing defects or deterioration is crucial [10].

When PD occurs, the type of defect can be identified through the Phase-Resolved Partial Discharge (PRPD) [11]. PRPD distinguishes defects by deriving their patterns through the combination of data on PD pulse amplitude q, phase Φ, and number of pulse n. Therefore, it is extremely important for monitoring the condition of cast-resin transformers and for the early detection of potential hazards. However, traditional analysis methods have the disadvantage of relying heavily on the experience of experts. Accordingly, there is a need for Artificial Intelligence (AI)-based systems that utilize rapidly advancing AI technologies to offer rapid processing and consistent diagnostic outcomes [12].

Among various AI technologies, the remarkable improvement in the performance of Convolutional Neural Network (CNN) has led to their widespread use in the field of image recognition and processing [13]. This is attributed to the capability of CNN to extract significant features from complex image data and to effectively learn the local characteristics of images, thereby enabling accurate classification. Recognizing and classifying partial discharge patterns are also crucial, as they can identify specific defects or prevent accidents in advance. Therefore, recent studies are increasingly focusing on applying CNNs for accurate recognition and classification of partial discharge patterns [14], [15], [16], [17], [18].

However, most AI systems, including CNN, have the black box problem, failing to provide clear interpretations for their predictions or decisions [19], [20]. This can cause serious issues in areas where sensitive or critical decisions are required. In the case of PD defect classification, if it is not clear on what criteria or basis the CNN model categorizes defects, its reliability decreases, and incorrect results can lead to serious accidents. Additionally, if the training process of the CNN model is not properly conducted, it becomes significantly challenging to analyze and resolve the causes of the errors. This becomes a significant obstacle in improving the model's performance and ensuring its reliability. To address this issue, research on the application of eXplainable Artificial Intelligence (XAI) is being conducted across various fields, aiming to make AI's decision-making process transparent, allowing users to comprehend and trust the decisions made [21], [22]. However, research on applying XAI in AI-based PD classification is currently inadequate. Hence, there is a demand for studies in XAI that allow users to comprehend the PD classification outcomes produced by CNN models, increase reliability through validation and analysis, and proactively prevent potential errors.

In this paper, we applied a CNN model, known for its excellence in image recognition, for PD classification in cast-resin transformers, and utilized the Gradient Weighted Class Activation Mapping (Grad-CAM) model among XAI techniques to propose a method for humans to understand the reasons for the results. The data used for training and validating the CNN model consists of artificial defect measurements under laboratory conditions and noise measured on a cast-resin transformer via UHF sensors. After training the CNN model with the PRPD pattern, the explanations for successful and failed prediction results were derived from XAI images. This approach suggested interpretability of the CNN model and high reliability in PD classification.



(a)

(b)

(c)

FIGURE 1. Noise measured using a UHF sensor installed on a cast-resin transformer (a) background noise or external noise, (b) and (c) UPS noise.

II. PRPD PATTERN MEASUREMENT PROCESS

A. NOISE MEASUREMENT

Noise measurement data was obtained through a UHF sensor installed on an operating cast-resin transformer, as shown in Fig. 1. Fig. 1(a) depicts patterns distributed across all phases. It is similar to the pattern of floating discharge, but there are differences depending on the phase. Therefore, it is estimated as background noise or introduced external noise. Fig. 1(b) and Fig. 1(c) show the noise from Uninterruptible Power Supply (UPS) measured by UHF sensors in cast-resin transformers. UPS is a device that provides stable power to critical loads even when the main supply is interrupted. The structure of the UPS is depicted in Fig. 2. It converts AC power to DC through a rectifier, supplying power to the batteries and the inverter. In case of an outage, the stored DC power in the batteries is transferred to the inverter side, converted to AC, and then supplied to the load. Fig. 3 illustrates the measured noise of a UPS intended for installation in a cogeneration power plant. Fig. 3(b) shows the noise measured during the rectifier's operation, while Fig. 3(c) presents the noise during the inverter's operation, which combines increased magnitude at certain phases with the noise of the rectifier. Consequently, it was confirmed that Fig. 1(b) and 1(c) represent the noise from the UPS's rectifier and inverter.

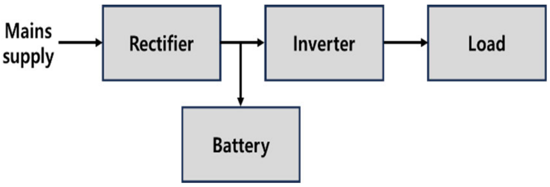


FIGURE 2. Block diagram of UPS.

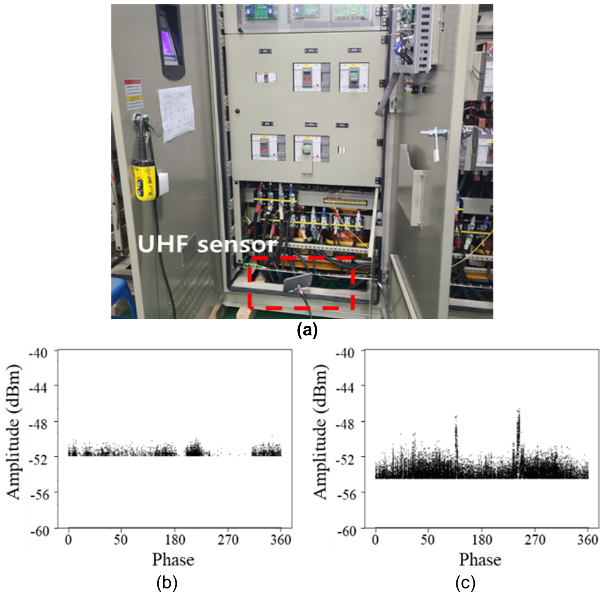


FIGURE 3. UPS noise measurements (a) UPS installed in cogeneration power plant, (b) rectifier noise and (c) inverter noise.

B. ARTIFICIAL DEFECT FABRICATION FOR PD MEASUREMENT

The artificial defects of the cast-resin transformer are depicted in Fig. 4. Fig. 4(a) was created as an artificial defect for surface discharge caused by cracks or contamination in epoxy. To measure surface discharge, epoxy with a 40 mm diameter was chosen as the dielectric material. In order to confirm the difference in PRPD pattern according to surface length, the diameters of the copper upper electrodes were set to 30mm and 20mm, respectively, resulting in surface distances of 5 mm and 10 mm. In the manufacturing process of cast-resin transformers, void discharge continuously occurs due to the imperfect removal of air bubbles or the formation of internal voids over extended periods of stress. This was manufactured as an artificial fault, as depicted in Fig. 4(b). To measure void discharge, a void was created between two electrodes. the internal void size was manufactured at 0.5Φ and 3Φ. During the process of filling molds with epoxy to produce void defects, there were cases where foreign materials were introduced, or the desired void shapes were not achieved. Therefore, 10 samples of each were produced. Fig. 4(c) represents a corona discharge. Corona discharge in cast-resin transformers occur more due to external factors such as protruding metallic particles or dust during manufacturing and installation, rather than within the dielectric material. Consequently, artificial defects were fabricated using a stainless-steel needle electrode with a curvature radius of 200 μm, considering the electric field concentration at external protruding electrodes. The distance between the needle electrode and ground is 15cm. Fig. 4(d) is a floating discharge, which has a very low probability of occurring in cast-resin transformers and does not need to be considered. However, as discharges suspected to be background noise resemble floating discharge, they were selected as defects to verify the classification and basis of the results after CNN training. The electrode, designed for floating discharge simulations and made of copper, was rounded at the bottom to an R10 curvature. The top part, with a 0.5 mm gap from the upper layer, was fabricated to a length of 10 mm.

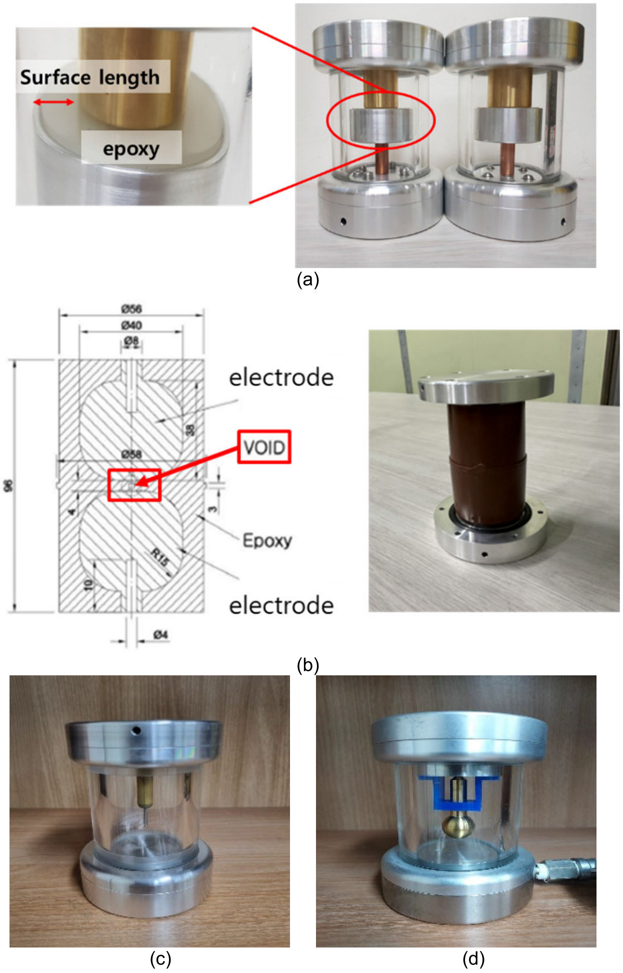
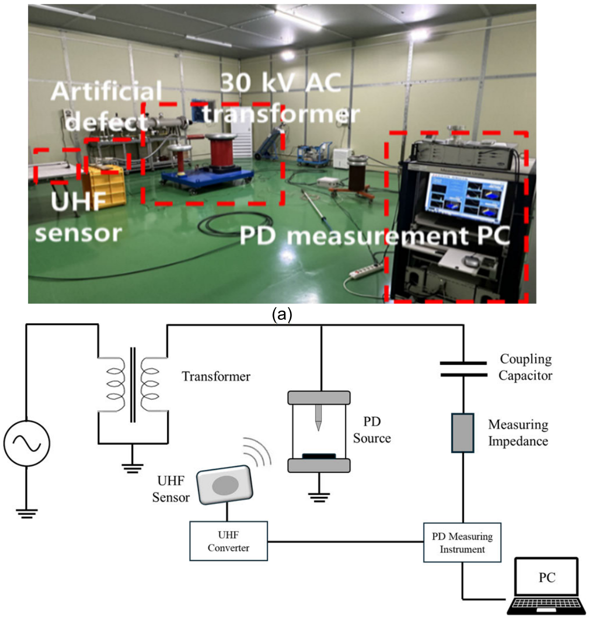


FIGURE 4. Artificial defect (a) surface discharge, (b) void discharge, (c) corona discharge and (d) floating discharge.

C. PD MEASUREMENT SYSTEM

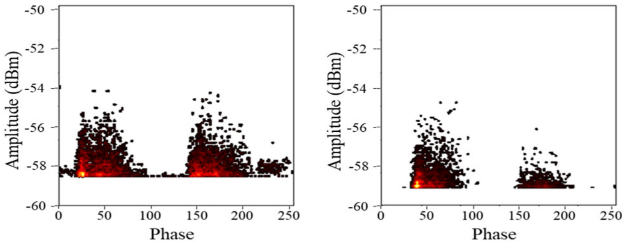
The PD measurement system is composed of a shielded room, AC power source, artificial defect, UHF sensors, and a PD measurement PC, as depicted in Fig. 5. To prevent the measurement of noise during PD testing, experiments were conducted in a shielded room. The AC power source can apply AC power of 60Hz up to 30 kV. The bandwidth of the UHF sensor is 300–800 MHz, and it is the same type



(a)

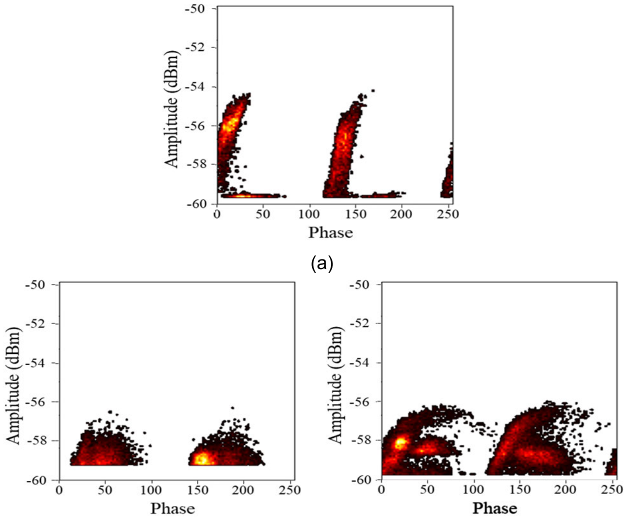
(b)

FIGURE 5. PD measurement system (a) photograph and (b) schematic diagram.



(a) (b)

FIGURE 6. Surface discharge pattern (a) 1.0mm and (b) 0.5mm.



(a)

(b) (c)

FIGURE 7. Void discharge pattern (a) 3Φ, (b) and (c) 0.5Φ.

of attachable sensor used for measuring noise in cast-resin transformers. The distance between the artificial defects and the UHF sensor was set to 50 cm, identical to the distance of the UHF sensor installed in the cast-resin transformer. The PD signal measured through the UHF sensor is saved to the PC. The PD measurement PC divides one cycle of 60Hz into 256 segments for storage, and these segments are accumulated to derive the PRPD pattern.

TABLE 1. PDIV and number of PRPD data.

Artificial defect	PDIV	Number of data
surface discharge (10mm)	5.6 kV	91
surface discharge (5mm)	4.8 kV	77
void discharge (3Φ)	12 kV	77
void discharge (0.5Φ)	9.7 kV	65
corona discharge	6.4 kV	220
floating discharge	3.4 kV	93
noise	-	52

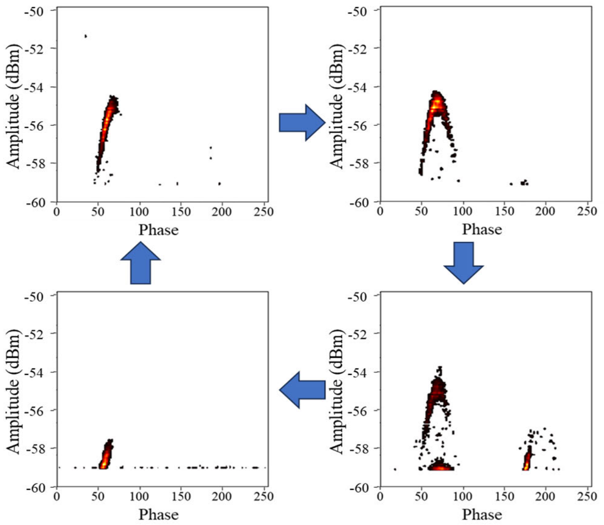


FIGURE 8. Corona discharge pattern change according to voltage application time.

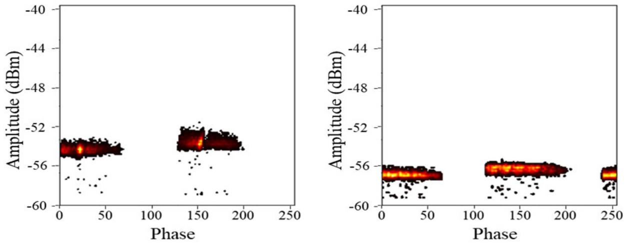


FIGURE 9. Floating discharge pattern.

D. PRPD PATTERNS

PD was measured for each defect, and the number of data obtained, and the Partial Discharge Inception Voltage (PDIV) are as presented in Table 1. Fig. 6(a) is a surface discharge pattern in which the upper electrode is small and the distance between the epoxy and the electrode is long, and Fig. 6(b) is a discharge pattern in which the distance is short. In Fig. 6(a), the PDIV is larger than in Fig. 6(b) because the distance between the electrode and the epoxy is longer. The PRPD

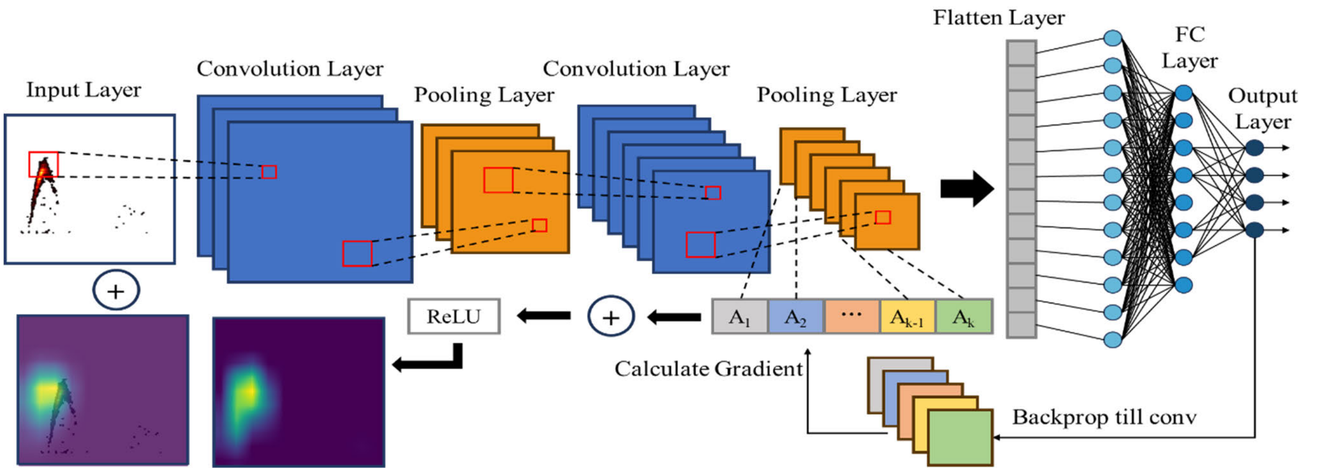


FIGURE 10. CNN and Grad-CAM architecture.

pattern was the same regardless of surface distance. The discharge was larger in the case of positive polarity, while it was smaller for negative polarity. As the discharge continued, the pattern size progressively decreased, and the discharge in the negative polarity almost vanished. Void discharge was measured in two sizes, 0.5Φ and 3Φ, as depicted in Fig. 7. 3Φ void had higher PDIV than 0.5Φ. PRPD patterns were different depending on the void size. Even when fabricated under identical conditions, the 0.5Φ defects were measured as depicted in Fig. 7(b) and 7(c). However, unlike corona and surface discharge, there was no change in pattern depending on the voltage application time. Fig. 8 shows the PRPD pattern of corona discharge and changes according to the applied voltage time. Initially measured starting at about 60 degrees phase, it appeared up to approximately 100 degrees, exhibiting a slanted pattern. After that, if the voltage is continuously applied, PD is measured up to a phase of about 120 degrees, and is drawn as a peak-shaped pattern. As the discharge continued, a small pattern was added to the bottom of the peak-shaped pattern. After that, dielectric breakdown occurred. After dielectric breakdown, the pattern was measured at a phase of approximately 60 degrees, and the above processes were repeated. The PRPD pattern of floating discharge is shown in Fig. 9. Discharges of consistent magnitude were measured in both the first and third quadrants. There was no difference in the patterns other than a slight shift in phase depending on the voltage application time.

III. GRAD-CAM APPLIED CNN MODEL

A. CNN AND GRAD-CAM ARCHITECTURE

For the classification of PRPD patterns in cast-resin transformers, a CNN and Grad-CAM were applied, with the architectures of both CNN and Grad-CAM depicted in Fig. 10. The structure of CNN is largely divided into feature extraction and Fully Connected Layer (FC). In the feature extraction stage, the process begins with the input image for classifying PRPD patterns. This is followed by convolution layers that extract features from the image using filters. Since maintaining the image size through to the FC layer would exponentially increase computational requirements, there is a pooling layer that appropriately reduces the size while emphasizing certain features. Once the features are extracted, a FC layer is necessary to classify what the image represents. The feature maps are arranged in a sequence identical to that of a conventional Deep Neural Network (DNN) and then fed into the FC layer. In this layer, every input neuron is connected to every neuron in the subsequent layer.

Grad-CAM is a technique that does not require the use of Global Average Pooling (GAP) and creates a heatmap by multiplying each feature map by the gradient. The superiority of Grad-CAM can be confirmed through the formulas and heatmap derivation process of both traditional CAM and grad- CAM.

$$L^{c}_{CAM}\,(i,j) = \sum\nolimits_{k} w^{c}_{k} f_{k}(i,j) \tag{1}$$

For the calculation of CAM, the flattening process following the last convolution layer is replaced with a GAP layer. In other words, the process involves calculating the average value of the feature map $\mathrm{f_k}$(i, j) from the last convolution layer, resulting in a single numerical output. The connection between the last convolution layer and the class is represented by weights w, and these are multiplied by $\mathrm{f_k}$(i, j) to produce k heatmaps. Following this, summing the heatmaps produces the resulting image of the CAM.

$$L^{c}_{Grad-CAM}\,(i,j) = ReLU(\sum\nolimits_{k} a^{c}_{k} f_{k}(i,j) \tag{2}$$

$$a^{c}_{k} = \frac{1}{Z}\sum\nolimits_{i}\sum\nolimits_{j}\frac{\partial S^{c}}{\partial f_{k}(i,j)}) \tag{3}$$

Through the formula, it was observed that a ReLU function was added, and the weights were replaced with gradients $\mathrm{a_k}$. This demonstrates that Grad-CAM, not incorporating a GAP layer, can be applied to a variety of CNN architectures. Furthermore, Grad-CAM can be applied not only to the final convolution layer but also to intermediate layers, enabling the observation of how the model processes information at various stages. Therefore, Grad-CAM was applied for XAI.

B. PROPOSED CNN MODEL

Table 2 presents the structure of the proposed CNN model. Each CNN block is composed of two CNN layers, two batch normalization layers, one maxpooling layer, and one dropout layer. The entire CNN model comprises three CNN blocks and a FC layer. The convolution layer is a key component of deep learning architectures specialized for visual data. It scans the entire PRPD pattern using filters to detect localized features, and a 3×3 size filter is applied to all convolution layers. Each filter has weights and a bias, with the 32 filters in the conv2d layer each containing 3×3 weights and a single bias. Weights and biases are updated through backpropagation and trained in the convolution layer. During the training process, each of the 32 filters acquires different values, indicating that each filter extracts different features of the image. Therefore, based on conv2, *number of parameters = (size of filter x number of input channels + 1) x number of filter*. According to the calculation formula, $(3 \times 3 \times 1 + 1) \times 32 = 320$.

Batch normalization, integrated within the neural network, adjusts the mean and variance together during training, preventing distorted distributions without separating them as an independent process. This enhances the training speed and helps to mitigate gradient loss. Batch normalization consists of four parameters: γ (scale), β (shift), moving mean, and moving variance. In the previous CNN model, each of the 32 filters is applied with four parameters, resulting in a total of 128 parameters to be trained. For Conv2_1, since the output channel of the preceding layer is 32, the number of parameters is calculated as $(3 \times 3 \times 32 + 1) \times 32 = 9248$.

The max pooling layer, along with the convolution layer, is a component of CNN model. A 2×2 window moves across the feature map in strides of two, reducing the image size and also simplifying computational complexity. While downsizing the image, the maximum value within each 2×2 window is selected. Through this, the image size is reduced to 58×58, forming 32 feature maps. This layer simply reduces the image size and does not require weights or training parameters.

Dropout is a technique to prevent overfitting by probabilistically deactivating some neurons, ensuring that the network does not become overly reliant on any specific neuron during training. In other words, some of the 32 feature maps are randomly set to zero, providing regularization to the network. Therefore, the parameter to train is 0. The process from the convolution layer to the dropout layer constitutes one CNN block, and this procedure was repeated two more times.

In a FC layer, every neuron in the current layer is connected to every neuron in the previous layer. In CNN, the output is derived in either 2D or 3D form, necessitating the use of a flatten layer to transform it into a 1D format. Since the output shape of the previous layer was $28 \times 28 \times 128$, passing through a flatten layer converts this into a format with 100,352 input neurons, calculated by multiplying the dimensions of the shape. Since it only alters the input shape, the number of trainable parameters in this process is zero.

TABLE 2. Constructed CNN model structure.

Layer info	Output shape	parameter
conv2d	(254, 254, 32)	320
batch_ normalization	(254, 254, 32)	128
conv2d_1	(254, 254, 32)	9248
batch_normalization_1	(252, 252, 32)	128
max_pooling2d	(126, 126, 32)	0
dropout	(126, 126, 32)	0
conv2d_2	(124, 124, 64)	18496
batch_normalization_2	(124, 124, 64)	256
conv2d_3	(122, 122, 64)	36928
batch_normalization_3	(122, 122, 64)	256
max_pooling2d_1	(61, 61, 64)	0
dropout_1	(61, 61, 64)	0
conv2d_4	(59, 59, 128)	73856
batch_normalization_4	(59, 59, 128)	512
conv2d_5	(57, 57, 128)	147584
batch_normalization_5	(57, 57, 128)	512
max_pooling2d_2	(28, 28, 128)	0
dropout_2	(28, 28, 128)	0
flatten	(100352)	0
dense	(512)	51380736
dropout_3	(512)	0
Dense_1	(5)	2565

Dense layers are used for learning and modeling complex relationships in data because they allow the network to learn different combinations of features from the input data. Therefore, all neurons from the previous layer are connected. In the proposed CNN structure, there are 100,352 input units fully connected to 512 output units. Each connection contains a weight, and the total number of parameters that need to be learned is _(input units*output units) + output units_, resulting in 51,380,736. Subsequently, a dropout layer and a dense layer were iteratively applied, setting the number of output units to five. This corresponds to the number of classes and represents the final stage of classifying the input data.

IV. RESULTS AND DISCUSSION

A. CLASS CLASSIFICATION RESULTS

Classification results were obtained by training the proposed CNN model with PRPD patterns measured through artificial defects and noise patterns measured from cast-resin transformers. The training data, comprising 675 samples, was divided into 70% for training and 30% for validation. This means 472 samples were utilized for training and 203 were used for validation. Before the training process, the parameters for epoch, learning rate, and batch size were set to 500, 0.0001, and 64, respectively. Upon completion of the training, the learning curve (Fig. 11) was configured to be displayed. The learning curve is a graph of validation accuracy and loss according to epochs. The accuracy and loss for training provide insights into how well the CNN model is adapting to the training data, allowing for the detection of underfitting. The training graph shows that from around epoch 25 onwards, the accuracy approaches 100%, and the loss nears almost 0%. Therefore, the training of the CNN

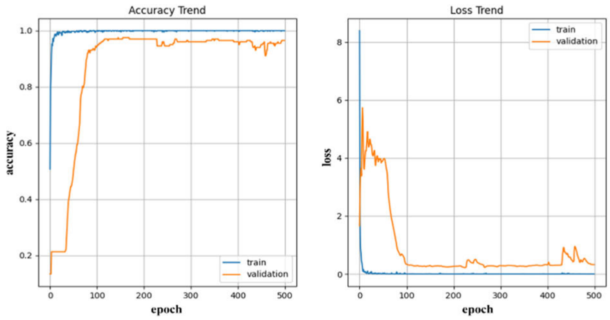


FIGURE 11. Accuracy and loss.

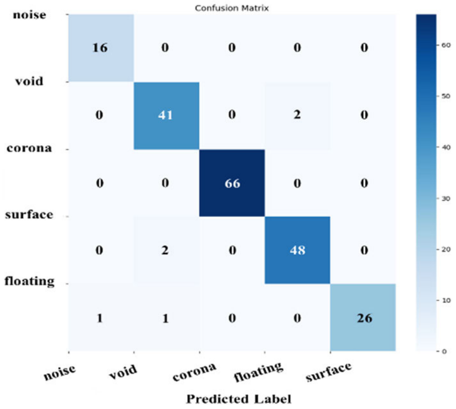
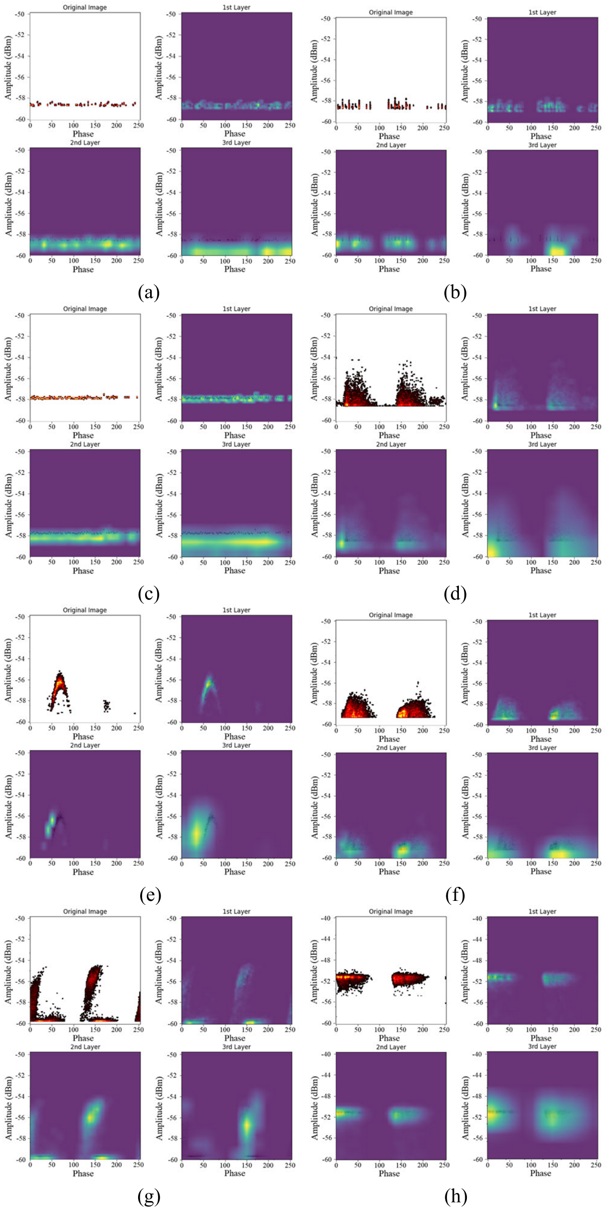


FIGURE 12. Confusion matrix.

model is progressing stably. Once the training is adequately completed, it is necessary to check the accuracy and loss for validation to assess the model's generalization ability. From epoch 100 onwards, the accuracy is around 97%, and the loss is close to 0%. Therefore, it is considered that there is no possibility of overfitting. The epoch with the lowest validation loss, number 226, was selected as the best model, and the classification results for the validation data were displayed as a confusion matrix, as shown in Fig. 12. Except for corona discharge and noise, the other types of discharge were misclassified by one or two instances each. Floating discharge was classified as void discharge and noise. Additionally, surface was classified as void discharge, and void was classified as surface discharge.



(a) (b) (c) (d) (e) (f) (g) (h)

FIGURE 13. Activation image via Grad-CAM (a) UPS rectifier noise, (b) UPS inverter noise, (c) background or introduced noise, (d) surface discharge, (e) corona discharge, (f) 3Φ void discharge, (g) 0.5Φ void discharge and (h) floating discharge.

B. ANALYSIS USING GRAD-CAM

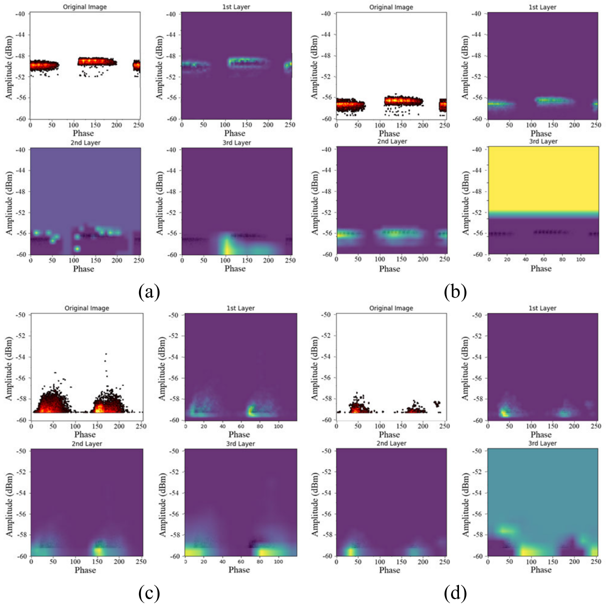
Since the CNN model cannot explain the reasoning behind its training results, only assumptions can be made about the causes of misclassification. However, applying Grad-CAM to the CNN model enables XAI. Therefore, using the proposed CNN model with Grad-CAM implementation, the basis for classification of PRPD and noise patterns for each defect was visualized and displayed in images, as illustrated in Fig. 13 and Table 3. Typically, the last convolution layer of a CNN model is visualized to determine which parts of an image were focused on for classifying a specific class. To understand how the proposed CNN model processes input data, intermediate layer were also visualized to identify which features were being learned. Fig. 13(a) and (b) respectively visualize UPS rectifier and inverter noise. Due to the absence of distinct features in the images, the 1st and 2nd layers appear very similar. In the 3rd layer, it begins to identify and display the characteristics of UPS inverter noise that enable class classification. However, the 3rd layer in Fig. 13(a) processes the data as if there are features across the entire phase. Fig. 13 (c) is background or introduced noise, and due to the monotonous image, there is no feature even if the convolution layer is deepened. Fig. 13(d) is a surface discharge, and the 1st layer focuses on features such as lines and corners of the image contours. As it progresses to the 2nd and 3rd layers, it starts to recognize patterns in the image and activates

TABLE 3. Summary of experiment results.

Artificial defect	Number of data	Accuracy	Grad-CAM analysis
surface discharge	168	96%	-Activation of important features increases with depth of layers
void discharge	142	97%	-Recognition of pattern differences based on void size
corona discharge	220	95%	-Activates more distinct features, such as during the positive half of the AC cycle
floating discharge	93	92%	-Activation of images in their original form
noise	52	94%	-With increasing depth of layers, the important features of UPS noise are highlighted

the important parts. Fig. 13(e) represents corona discharge, where difference in visualization can be seen in each convolution layer, similar to surface discharge. Small patterns were detected in the third quadrant of the input image. The 1st and 2nd layers show these patterns with a lighter visualization. However, in the 3rd layer, which detects higher-level features, the patterns measured at negative polarity are not considered significant. Consequently, only the positive polarity corona patterns are activated with bright colors. Fig. 13(f) and (g) represent void discharges of 0.5Φ and 3Φ sizes, respectively. The void discharge demonstrates more abstract features as the convolution layers become deeper. For the 3Φ size void, the pattern in the third quadrant is brighter than the pattern in the first quadrant. This indicates that the patterns in the third quadrant have a significant impact on the decision for this class. Fig. 13(f) is a floating discharge, visualized almost identically to the original image.

An analysis was conducted on the misclassified cases from the validation results using activated Grad-CAM images, as shown in Fig. 14. Fig. 14(a) and (b) illustrate the reasons for misclassifying floating discharge as void discharge and noise through images respectively. In the 2nd layer, except for a few dots, the focus was primarily on the background, and the activated Grad-CAM image in the third quadrant of the 3rd layer showed similarities to void discharge rather than the original floating discharge, leading to the misclassification. In the case of floating discharge classified as noise, the activated image in the 3rd layer was focused on the background. This indicates that the model did not reflect the characteristics of the floating discharge pattern, instead focusing on incorrect features. Although the colors are different, the bisected image resembles the 3rd layer of Fig. 13(b). Consequently, it was classified as noise. The reason for misclassifying a void discharge as a surface discharge was visualized and is shown in Fig. 14(c). There are vertical dots in the pattern of the third quadrant. In the 3rd layer, these vertically aligned dots have a low activation value, resulting in a lighter color representation. This was misclassified due to its similarity to the image of surface discharge. Fig. 14(d) shows the activated image of surface discharge. In the 3rd layer, the focus on the background creates a shape similar to void discharge, leading to incorrect class classification.



(a)
(b)
(c)
(d)

FIGURE 14. Analysis of misclassification causes through Grad-CAM (a) and (b) floating discharge, (c) void discharge and (d) surface discharge.

TABLE 4. Summary of image focused on the background.

Artificial defect	Observations through Grad-CAM
surface discharge	-Surface discharge shows characteristics of shooting upwards, and the convex-shaped activation image is produced due to larger y-axis values of the central pattern.
noise	-Floating discharge, suspected of resembling noise patterns, was also selected as defects to verify the classification and basis of results after CNN training.

C. DISCUSSION

The reasons for the class classification results of the CNN model were analyzed through Grad-CAM activation images. There was an error in class classification due to focusing on the background rather than the pattern of the image. If many data points had focused on the background rather than the features of the image, it would have been necessary to modify the structure or parameters of the CNN model. However, considering that only 10 out of the total data set focused on the background, it can be inferred that the CNN model is robust. As shown in Fig. 15 and Table 4, correct class classification was achieved even when surface and floating discharges were focused on the background. This is because surface discharges exhibit characteristics of shooting upwards, and floating discharges resulted in convex-shaped activation images due to the y-axis of the central pattern having larger values. Therefore, it was observed that even if the activation values are inverted in the image output, correct class classification is possible as long as the patterns are similar.

In cases of discharge estimated to be background or introduced noise, the 3rd layer shows activation values across the entire phase, unlike the original image. Therefore, a class imbalance exists, and it is believed to indicative of a lack of

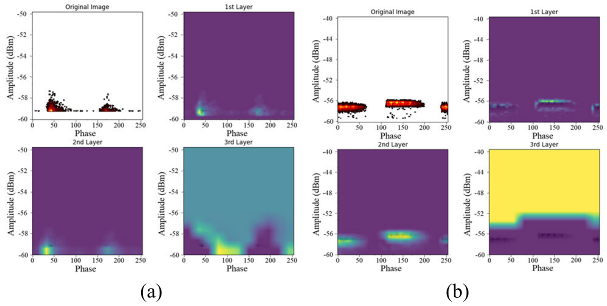


FIGURE 15. Accurate classification was achieved even with images focused on the background (a) surface discharge and (b) noise.

training data. There is a possibility that the activation images were output across the entire phase in the 3rd layer because sufficient training for noise did not occur. Alternatively, as the convolution layers become deeper, they include more abstract information, which could have led to results similar to those in Fig. 13(b). Therefore, it is necessary to obtain more noise data for additional validation. In the case of void discharge, it was misclassified as surface discharge due to the presence of vertically aligned dots. In the 99 data samples used for training, there were no such vertically aligned dots. Thus, it is presumed that these were noise infiltrated during the measurement of void discharge. To build a robust CNN model, it seems necessary to increase the number of data samples and include training for cases with partially noise introduced.

Previous studies focused on model architecture and performance optimization to perform partial discharge defect classification. Therefore, while they achieved improvements in computational efficiency and high accuracy in classifying PRPD patterns, the importance of model prediction transparency and user trust has been overlooked. This paper not only achieves classification accuracy comparable to previous studies but also integrates Grad-CAM with CNN to enhance user understanding. In other words, it is possible to deeply analyze the class classification results. This demonstrates that XAI is essential when integrating CNN models into diagnostic systems, with its applications being as follows.

1) The reason for class classification of the CNN model can be confirmed through visualization.
2) Analysis of incorrectly predicted outcomes allows for identification of error sources, facilitating the development of improved model architectures.
3) The model's reliability is enhanced through its verifiability by users.
4) Understanding the workings of the CNN model becomes feasible, offering valuable feedback for performance enhancement.

V. CONCLUSION

For the purpose of XAI, Grad-CAM was applied to the CNN model to classify PRPD patterns and verify the basis of the results. The model was trained on partial discharges measured from noise in cast-resin transformers and artificial defects. After training, the reasons for the successes and failures in class classification were analyzed using the images produced by Grad-CAM. The results are as follows.

- The validation results of the CNN indicated an accuracy of approximately 97%, showing no signs of overfitting.
- Reasons for successes and failures in the validation data were derived and analyzed using Grad-CAM images.
- It was identified that misclassifications occurred when activation images focused on the background or when noise introduced with void discharges.
- Securing a sufficient quantity of data is essential for the construction of a robust model.
- It is anticipated that implementing a CNN model equipped with Grad-CAM in diagnostic systems for XAI will enhance reliability compared to traditional AI models.

REFERENCES

[1] T. Nunn, "A comparison of liquid-filled and dry-type transformer technologies," in *Proc. IEEE-IAS/PCA Cement Ind. Tech. Conf. Rec.*, Sep. 2000, pp. 105–112.

[2] M. Eslamian, B. Vahidi, and A. Eslamian, "Thermal analysis of cast-resin dry-type transformers," *Energy Convers. Manage.*, vol. 52, no. 7, pp. 2479–2488, Jul. 2011.

[3] D. Azizian, M. Bigdeli, and J. Faiz, "Design optimization of cast-resin transformer using nature-inspired algorithms," *Arabian J. Sci. Eng.*, vol. 41, no. 9, pp. 3491–3500, Sep. 2016.

[4] P. K. Sen, "Application guidelines for dry-type distribution power transformers," in *Proc. IEEE Tech. Conf. Ind. Commercial Power Syst.*, Aug. 2003, pp. 105–110.

[5] Y.-L. Tan, "Damage of a distribution transformer due to through-fault currents: An electrical forensics viewpoint," *IEEE Trans. Ind. Appl.*, vol. 38, no. 1, pp. 29–34, Jan. 2002, doi: 10.1109/28.980341.

[6] F. Guastavino, E. Torello, A. Ratto, A. Dardano, M. Secci, F. Ferraro, and D. Pistone, "Diagnosis of common defects inside cast resin current transformers by digital partial discharges acquisition," in *Proc. 20th Int. Conf. Electr. Mach.*, Marseille, France, Sep. 2012, pp. 1647–1652, doi: 10.1109/ICElMach.2012.6350101.

[7] S. Makki, M. Kozako, M. Hikita, K. Iida, T. Umemura, S. Nakamura, Y. Nakamura, T. Hirose, T. Maeda, and M. Higashiyama, "Effect of X-ray irradiation on partial discharge inception and extinction voltage characteristics of cast resin transformer insulation system," in *Proc. IEEE Int. Conf. Solid Dielectr. (ICSD)*, Bologna, Italy, Jun. 2013, pp. 79–82.

[8] M. Bagheri, A. Subramaniam, S. Bhandari, S. Chandar, S. Nadarajan, A. K. Gupta, and S. K. Panda, "Dry-type transformer aging factor in various loads and environmental conditions," in *Proc. ITCE'15*, 2015, pp. 1–6.

[9] U. Kaltenborn and T. Miessler, "On-site testing and PD diagnosis of cast-resin distribution transformers," in *Proc. Int. Conf. Diag. Electr. Eng.*, Pilsen, Czech Republic, Sep. 2020, pp. 1–5, doi: 10.1109/Diagnostika49114.2020.9214595.

[10] Y.-W. Tang, C.-C. Su, C.-C. Tai, and J.-F. Chen, "Cast-resin dry-type transformer partial discharge signal analysis using spectral correlated empirical mode decomposition method," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, Minneapolis, MN, USA, May 2013, pp. 523–527, doi: 10.1109/I2MTC.2013.6555472.

[11] H. Illias, T. Soon Yuan, A. H. A. Bakar, H. Mokhlis, G. Chen, and P. L. Lewin, "Partial discharge patterns in high voltage insulation," in *Proc. IEEE Int. Conf. Power Energy*, Kota Kinabalu, Malaysia, Dec. 2012, pp. 750–755, doi: 10.1109/PECON.2012.6450316.

[12] A. Mas'ud, R. Albarracín, J. Ardila-Rey, F. Muhammad-Sukki, H. Illias, N. Bani, and A. Munir, "Artificial neural network application for partial discharge recognition: Survey and future directions," *Energies*, vol. 9, no. 8, p. 574, Jul. 2016, doi: 10.3390/en9080574.

[13] C. Iorga and V.-E. Neagoe, "A deep CNN approach with transfer learning for image recognition," in *Proc. 11th Int. Conf. Electron., Comput. Artif. Intell. (ECAI)*, Jun. 2019, pp. 1–6, doi: 10.1109/ECAI46879.2019.9042173.

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

[15] M. Florkowski, "Classification of partial discharge images using deep convolutional neural networks," *Energies*, vol. 13, no. 20, p. 5496, Oct. 2020, doi: 10.3390/en13205496.

[16] T. Liu, J. Yan, Y. Wang, Y. Xu, and Y. Zhao, "GIS partial discharge pattern recognition based on a novel convolutional neural networks and long short-term memory," *Entropy*, vol. 23, no. 6, p. 774, Jun. 2021, doi: 10.3390/e23060774.

[17] Y. Liu, M. Hu, Q. Dai, H. Le, and Y. Liu, "Online recognition method of partial discharge pattern for transformer bushings based on small sample ultra-micro-CNN network," *AIP Adv.*, vol. 11, no. 4, pp. 1–12, Apr. 2021, doi: 10.1063/5.0047481.

[18] S. Mantach, P. Gill, D. R. Oliver, A. Ashraf, and B. Kordi, "An interpretable CNN model for classification of partial discharge waveforms in 3D-printed dielectric samples with different void sizes," *Neural Comput. Appl.*, vol. 34, no. 14, pp. 11739–11750, Jul. 2022.

[19] A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, Jun. 2020.

[20] A. M. Antoniadi, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, and C. Mooney, "Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: A systematic review," *Appl. Sci.*, vol. 11, no. 11, p. 5088, May 2021.

[21] W. Saeed and C. Omlin, "Explainable AI (XAI): A systematic meta-survey of current challenges and future opportunities," *Knowl.-Based Syst.*, vol. 263, Mar. 2023, Art. no. 110273.

[22] B. H. M. van der Velden, H. J. Kuijf, K. G. A. Gilhuijs, and M. A. Viergever, "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102470.

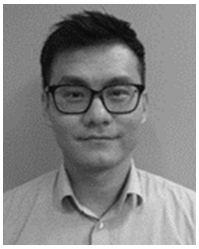
HO-SEUNG KIM received the B.S. degree in physics from Changwon National University, Changwon, South Korea, in 2019. He is currently pursuing the Ph.D. degree with the Department of Electronic Engineering, Hanyang University, Ansan, South Korea. His research interests include insulation materials, partial discharge, and HVDC.

JIHO JUNG received the B.S. degree in electrical engineering from Hanyang University, Ansan, in 2022, where he is currently pursuing the M.S. degree with the Department of Electronic Engineering. His research interests include high voltage engineering and HVDC.

RYUL HWANG received the B.S. degree in electrical engineering from Hoseo University, Asan, South Korea, in 2002. He is currently pursuing the Ph.D. degree in electrical engineering with Hanyang University, Ansan, South Korea. His research interests include insulation materials, high-voltage discharge, insulation design, high-voltage power equipment, and partial discharge.

SEONG-CHAN PARK received the B.S., M.S., and Ph.D. degrees from the Department of Electrical and Computer Engineering, Seoul National University, Seoul, South Korea, in 2009, 2011, and 2016, respectively. He was a Research Engineer with LS Industrial Systems Company Ltd., South Korea. In 2021, he joined the Electricity Technology Team, Samsung Electronics Company Ltd., Hwaseong, South Korea. His research interests include AI applications in power systems and deregulated power markets.

SEUNG-JAE LEE received the B.S. and M.S. degrees from the Department of Electrical and Electronic Engineering, Yonsei University, Seoul, South Korea. He joined the Infra Technology Innovation Team, Samsung Electronics Company Ltd., Hwaseong, South Korea. His research interests include cable diagnosis by applying signal processing and structural health monitoring by applying piezoelectric sensors.



GYU-TAE KIM received the B.S. degree from the Department of Electrical Engineering, Chungbuk National University, and the M.S. and Ph.D. degrees from the Department of Electronic and Electrical Engineering, Pohang University of Science and Technology, South Korea. He joined the Infra Technology Innovation Team, Samsung Electronics Company Ltd., Hwaseong, South Korea. His research interests include power electronics and diagnostics of electrical equipment faults.

BANG-WOOK LEE (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees from the Department of Electrical Engineering, Hanyang University, Seoul, South Korea, in 1991, 1993, and 1998, respectively. He was a Senior Research Engineer with LS Industrial Systems Company Ltd., South Korea. In 2008, he joined the Department of Electronic Engineering, Hanyang University, Ansan, South Korea, where he is currently a Professor. His research interests include HVDC protection systems, high voltage insulation, renewable energies, the development of electrical equipment, and transmission line structures for HVDC and HVAC power systems. He is a member of the HVDC Research Committee of KIEE, the Power Cable Experts Committee of the Korean Agency for Technology and Standards, and CIGRE.