Convolutional-Neural-Network-Based Partial Discharge Diagnosis for Power Transformer Using UHF Sensor

THE-DUONG DO[1], **VO-NGUYEN TUYET-DOAN**[1], **YONG-SUNG CHO**[2], **JONG-HO SUN**[2], **AND YONG-HWA KIM**[1], **(Member, IEEE)**

[1]Department of Electronic Engineering, Myongji University, Yongin 17058, South Korea
[2]Advanced Power Apparatus Research Center, Korea Electrotechnology Research Institute, Changwon 51543, South Korea

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

ABSTRACT Given the enormous capital value of power transformers and their integral role in the electricity network, increasing attention has been given to diagnostic and monitoring tools as a safety precaution measure to evaluate the internal condition of transformers. This study overcomes the fault diagnosis problem of power transformers using an ultra high frequency drain valve sensor. A convolutional neural network (CNN) is proposed to classify six types of discharge defects in power transformers. The proposed model utilizes the phase–amplitude response from a phase-resolved partial discharge (PRPD) signal to reduce the input size. The performance of the proposed method is verified through PRPD experiments using artificial cells. The experimental results indicate that the classification performance of the proposed method is significantly better than those of conventional algorithms, such as linear and nonlinear support vector machines and feedforward neural networks, at 18.78%, 10.95%, and 8.76%, respectively. In addition, a comparison with the different representations of the data leads to the observation that the proposed CNN using a PA response provides a higher accuracy than that using sequence data at 1.46%.

INDEX TERMS Partial discharge (PD), fault diagnosis, power transformer, and convolutional neural network (CNN).

I. INTRODUCTION

Over the past several decades, with the development of the economy and the continuous advancement of society, the global energy industry has witnessed rapid growth, and electric energy demand has become increasingly vigorous. As the primary origin for changing the voltage, the power transformer, one of the most important components for maintaining stable operational conditions of a power system, assumes an inevitable role in the network of transmission and dispensing systems [1]–[3]. Working under factual conditions such as the process of manufacturing, installation, maintenance, and a prolonged period of operation, the transformer is unavoidably subjected to external factors such as power, apparatus, and heat. This may lead to the degradation of insulation over a

period by gradual erosion in the winding, generating fractional release separator phenomena (partial discharge (PD)), compromising the stable operation condition of the entire system [4], [5]. In addition, if they are not detected at an early stage, PD failures will progressively evolve, and the transformer insulation will increasingly deteriorate. Ultimately, PD will develop into a discharge breakdown or spark discharge, which may result in full insulation deterioration and tremendous economic losses [6], [7]. As a result, monitoring and diagnostic techniques that assess the stable reliability, predict the presence, and identify the types of PD insulation defects in the transformer precisely in a timely manner are of great significance for enhancing the operational reliability of power transformers and maintaining the safe operation of the power grid and the electricity supply [8]–[10].

Electrical measurements based on IEC 60270 are widely used for power transformer testing during routine factory

---

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

tests [11]. However, it is difficult to measure and analyze PD owing to electromagnetic noise in the substations. This has led to numerous research endeavors in many nations to look for alternative methods of PD diagnosis. Recently, to identify PD signals, measurements using conventional-pulse current, high-frequency current, ultrahigh frequency (UHF), acoustic emission (AE), and dissolved gas analysis (DGA) have been gaining popularity in the monitoring and evaluation of power transformers [5]. Among these methods, the pulse current and DGA methods are not suitable for providing locations for PD sources, and the acoustic method has the disadvantage of diagnosing faults inside the power transformer [12], [13]. Therefore, the UHF method is widely used for PD diagnosis because of its high sensitivity and robustness to noise [13]. The recognition of UHF signals can provide extremely valuable information in analyzing the internal state, allowing one not only to identify faults in the device but also to determine what kind of faults they are, where they are, and how dangerous they are.

With the progression of signal processing techniques, numerous studies have focused on applying artificial intelligence and optimization techniques for the purpose of monitoring the condition of power system components as well as improving and enhancing the accuracy of diagnosing the fault, consisting of transformer fault diagnosis. Among the various detection techniques, UHF has shown promising results in PD recognition and classification [14]–[17]. Classical PD types have been measured using the UHF antenna, and then a concept based on zero-span was proposed for PD classification in [14]. An enhancement of the diagnostic process with more than one source of PD or interference was presented in [15]. In that study, wavelet-based time—-frequency analysis was applied to find suitable features in PD data to distinguish between PD sources at different locations. This technique is based on calculating the similarity between pairs of signals that have been transformed into the time—-frequency domain. For on-site PD, reliability of the diagnosis has been achieved by utilizing a redundant diagnosis strategy-based system [16].

With the recent and rapid growth of technology, machine learning algorithms, such as artificial neural networks (ANNs), support vector machines (SVMs), and decision trees [17]–[23] have been introduced in the literature as promising techniques for a fault diagnosis of a transformer. Of them, the combination of signal processing techniques and SVMs, which can handle the main issues of a ''dimensionality disaster,'' ''over-fitting,'' and local minimum point, has gained increasing prevalence [18]. To classify the PD types in a power transformer, feature extraction of the signals has been accomplished based on a wavelet analysis, and an improved bagging algorithm with a backpropagation neural network and an SVM were then used for the classification task [17]. However, it is challenging to accurately choose the appropriate model parameters, including the kernel function, which has a significant influence on the classification performance of an SVM [19], [20]. The fuzzy theory has a simple structure

and achieves a fast diagnosis; however, its learning ability is insufficient, and it does not have the ability to take advantage of previous diagnosis results [21]. For a clear description of the uncertainties and to overcome the limitation of automatically adjusting a fuzzy rule diagnosis, one of the most widely used artificial intelligence methods, an ANN with superior learning capabilities, which generates an efficient structured network with weight vectors instead of fault diagnosis rules, has been introduced for a transformer fault prediction [22], [23]. Although an ANN has strong self-learning and a parallel processing capability to learn from the training data directly, tackling nonlinear relationships and generalized solutions for a new dataset, the convergence of the model is quite slow, oscillation occurs at times, and it easily falls into a local optima [18]. In addition, to acquire the best performance of the network, the selection of parameters of the ANN model, such as the number of hidden layers and the hidden neurons in each layer, must be properly considered. Despite the cooperation of features and the fact that such algorithms have achieved a substantial level of success, it has been perceived that such conventional methods have reached a bottleneck in their development, which restricts the further improvement of the pattern recognition accuracy.

To address these problems, various studies have been conducted using deep learning, which has been recognized as a mainstream of AI because deep learning has shown a much higher accuracy than other machine learning approaches. In particular, convolutional neural networks (CNNs), which have arguably the most widespread deep learning architecture, have increasingly shown prominent results for a variety of computer vision problems such as image classification, segmentation, detection, and video tracking, and have achieved prevailing success and an exemplary performance in the field of pattern classification and retrieval related tasks in recent years [24], [25]. It must first be recognized that the principal dominance of a CNN in comparison to its predecessors is that it automatically and adaptively learns representations by exploiting spatial or temporal correlations in the data based on the use of multiple feature extraction stages and detects the essential characteristics without any human surveillance. In comparison to other deep learning techniques, the complexity of a CNN and the difficulty of its training are greatly diminished through the sharing of parameters and local connections, which additionally decrease the risk of over-fitting [26]. In addition, constructing a network with a deep architecture utilizing a CNN is relatively easy. Similar to many other deep neural networks, the topology of a CNN is separated into multiple learning stages from the input to the output layers composed of a combination of multiple hidden layers, such as convolution layers, pooling layers, a dropout, and fully connected layers in the middle. Each layer has numerous filters, kernels, or neurons that respond to distinctive combinations of inputs from the prior layers, accomplish a convolution, and take its place with an optional non-linearity [27]. In addition to their sample efficiency in achieving accurate models, CNNs tend to be computationally

efficient, both because they require fewer parameters than fully connected architectures and because convolutions are easy to parallelize across GPU cores.

The main objective of this work is to develop a convolutional neural network (CNN) model to address the essential requirements of PD source classification in power transformers. We utilize the characteristics of CNNs to classify the incipient defects of power transformers using the phase–amplitude (PA) response of the PD signals as the input of the CNNs. The structure of the proposed CNN model in our experiment includes convolutional layers and max-pooling layers with the main task of performing feature extraction of the PRPD signal. The dropout and batch normalization hold the responsibility of preventing overfitting. In addition, fully connected layers are applied in our model as a function of mapping the extracted features into the final output, where a classification layer is employed to recognize different insulation faults for the power transformer. The main contributions of this article are as follows.

- A UHF drain valve sensor was used to capture the waves emitted by six types of discharge defects that may occur inside the power transformer [28], [29]. The performance of the proposed method was verified through experiments using artificial cells for the power transformer. In addition, noise is regarded as a normal state to analyze the influence of noise in PD classification.
- The PA response is obtained from the PRPD to decrease the size of the input matrix for the CNN. The PA response has different characteristics for each failure in the power transformer. The proposed CNN with the PA response has better classification performance than the CNN method using PRPDs and achieved a classification performance of almost 100%.
- The proposed CNN has higher classification accuracy than previous classification methods such as linear SVM, nonlinear SVM, and FNN. This is because the proposed CNN has the advantage of feature extraction based on the PA response.

The remainder of this article is organized as follows. In Section II, the description of PRPDs and noise measurements for power transformers is presented. Then, the proposed CNN-based classification method, including the input representation and the framework for CNN, is presented in Section III. Performance evaluations are presented in Section IV. We also compare the performance of the proposed method with other classification methods in this section. Finally, Section V concludes this article. In addition, the acronyms used in this article are listed in Table 1.

II. PRPD MEASUREMENT USING UHF SENSOR
In this section, we present our experimental setup and experimental results for PRPDs in order to investigate the assessment of the PD characteristics in the power transformer. Artificial cells are modeled for six types of faults in power transformers, and noise measurements are also conducted.
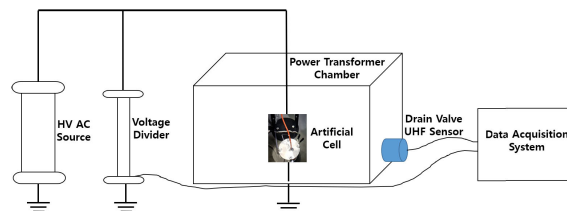

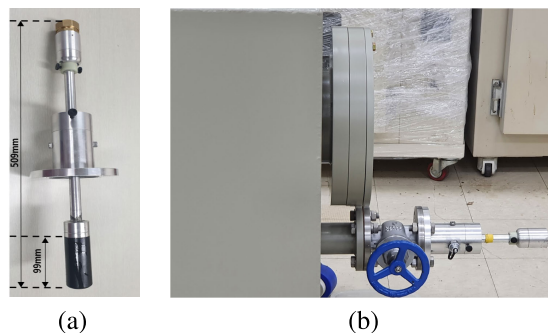**FIGURE 1.** Block diagram of experimental system for power transformer.

(a)   (b)

FIGURE 2. UHF drain valve sensor for oil-insulated transformers: (a) sensor external shape and (b) sensor installation.

TABLE 1. List of acronyms and abbreviations.

Acronym	Description
2D	Two-dimensional
AC	Alternating current
ADC	Analog-to-digital converter
AE	Acoustic emission
CNN	Convolutional neural network
DAS	Data acquisition system
DGA	Dissolved gas analysis
FC	Fully connected
FNN	Feedforward neural network
PA	Phase amplitude
PD	Partial discharge
PRPD	Phase-resolved partial discharge
ReLU	Rectified linear unit
SVM	Support vector machine
t-SNE	t-distributed stochastic neighbor embedding
UHF	Ultra-high frequency

A. EXPERIMENTAL SYSTEM
Fig. 1 shows a pictorial block diagram of the experimental test system, which is composed of a power transformer chamber, UHF drain valve sensor, and data acquisition system (DAS).

The UHF drain valve sensor using a monopole-type broadband antenna is designed to be suitable for the UHF frequency band of PDs generated from a power transformer and installed in the oil-filled power transformer chamber, as shown in Fig. 2. As shown in Fig. 3, the DAS comprises an amplifier, a log detector, an 8-bit analog-to-digital converter (ADC), and a personal computer (PC). The amplifier has a gain of 45 dB within the frequency range of 300 MHz to 1.8 GHz. The log
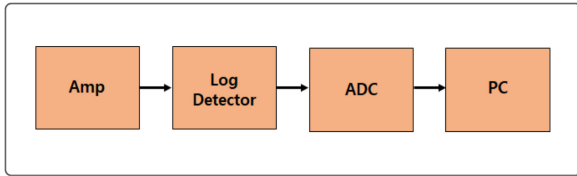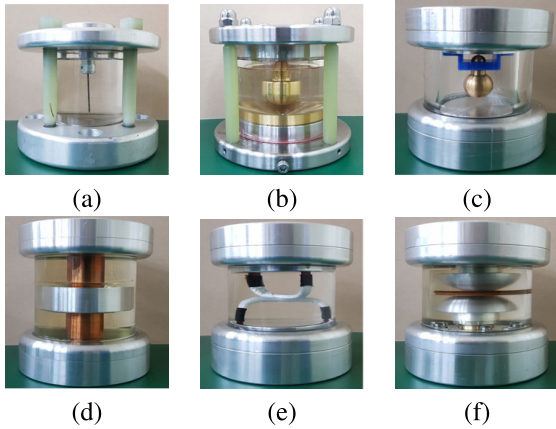
FIGURE 3. Block diagram of DAS.

FIGURE 4. Artificial cells: (a) protruding electrode, (b) particle, (c) floating, (d) surface, (e) bad contact between windings, and (f) void PDs.

detector consists of a logarithmic amplifier and a peak detector, where the logarithmic amplifier is used for the dynamic range compression, and the peak detector is used to capture the maximum values of the UHF PD pulses [30]. After the peak detector, the DAS uses the ADC with $1024 \times f_m$ samples per second, where $f_m = 60$ Hz is the power frequency. The maximum value is then captured at every 8 samples in the DAS, and $P = 128$ samples in each power cycle were used for the PRPD measurements.

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

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

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

B. PRPD MEASUREMENTS

PRPD measurements according to the form and characteristics of PD signals were performed in artificial cells filled with oil. Fig. 4 shows six artificial cells used for simulating various types of PD insulation defects in power transformers such as protruding electrode discharge, floating discharge, free particle discharge, surface discharge, bad contact between windings, and void PDs [31]. Moreover, the voltage divider and high-voltage AC source for generating PDs in the artificial cells were applied in the PRPD measurements. The test voltages for artificial cells relevant to each type in the experiment are given in Table 2.

TABLE 2. Artificial cells under different voltage levels.

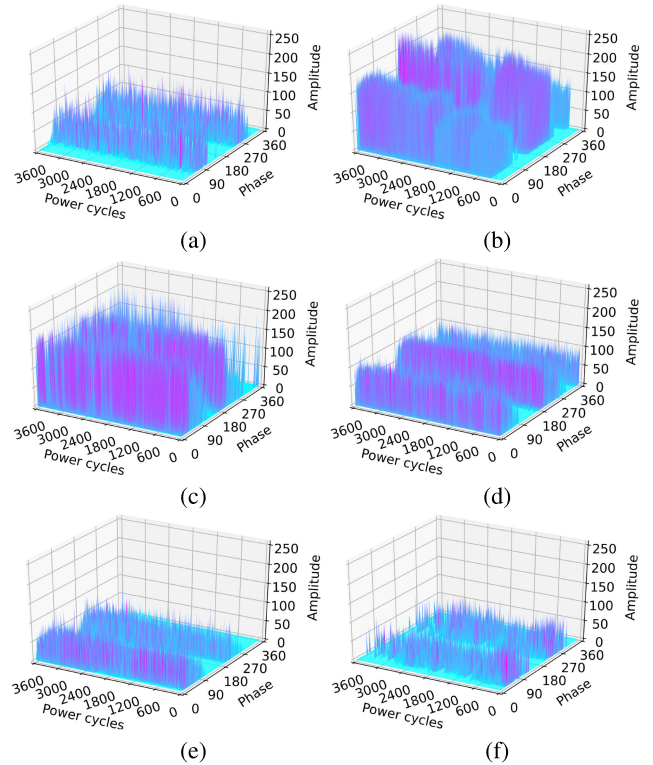| No.	Artificial cells	Voltage (kv)
1	Protruding electrode	14.3
2	Particle	23.0
3	Floating	10.6
4	Surface	14.0
5	Turn-to-turn	4.0
6	Void	10.5

FIGURE 5. Sequential phase-resolved PDs (PRPDs) for six fault types in power transformer: (a) protruding electrode, (b) floating, (c) particle, (d) surface, (e) turn-to-turn, and (f) void.

Fig. 5 presents sequential data for PRPDs under six types of insulation defects with 3600 power cycles in the six artificial cells using the UHF drain valve sensor. As can be seen from Fig. 5a that the discharge pulses of the protruding electrode are observed separately at both a positive and negative half-cycle near 90° and 270°, which have a distribution similar to that of void discharge. However, the pulses of the void are sparser. For the floating and particle, there is an extremely condensed density of evident discharge pulses across all bands at different intensity ranges, as shown in Figs. 5b and 5c, in which the amplitude of floating PDs reaches a record-high of 250. Meanwhile, the discharge pulses of the turn to turn are mainly distributed in the first quadrant. In addition, it can be seen that some sparse pulses emerged in the third quadrant near 270°. Fig. 5d shows that the presence of surface PDs with high amplitude is found massively in the regions
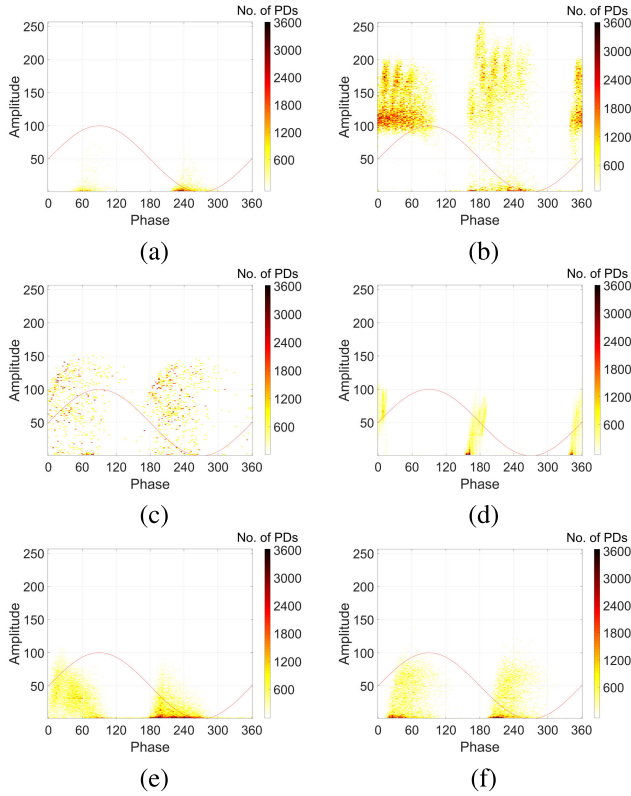
FIGURE 6. PA responses of PRPDs for six fault types in power transformer: (a) protruding electrode, (b) floating, (c) particle, (d) surface, (e) turn-to-turn, and (f) void.

FIGURE 7. Example of noise measurements: (a) phase-resolved PD (PRPD) and (b) PA response.

TABLE 3. Detail structure of proposed CNN model.

Layer type	Kernel size/stride	Kernel number	Output size	Padding
Conv+ReLU	$5 \times 5/2$	128	$64 \times 128 \times 128$	same
Max-pooling	2×2	-	$32 \times 64 \times 128$	valid
Dropout	0.5	-	$32 \times 64 \times 128$	-
Batch norm	-	-	$32 \times 64 \times 128$	-
Conv+ReLU	$3 \times 3/2$	64	$16 \times 32 \times 64$	same
Max-pooling	2×2	-	$8 \times 16 \times 64$	valid
Dropout	0.5	-	$8 \times 16 \times 64$	
Batch norm	-	-	$8 \times 16 \times 64$	-
Conv+ReLU	$3 \times 3/2$	16	$4 \times 8 \times 64$	same
Dropout	0.5	-	$8 \times 8 \times 64$	
Batch norm	-	-	$4 \times 8 \times 64$	-
Conv+ReLU	$3 \times 3/2$	32	$23 \times 4 \times 32$	same
Batch norm	-	-	$2 \times 4 \times 32$	-
Conv+ReLU	$3 \times 3/2$	12	$1 \times 2 \times 12$	same
Flatten	-	-	24	-
Fully connected	-	-	256	-
Softmax layer	-	-	7	-

around 0° - 90° in the positive band, and from 180° to 300° in the other band, and a smaller part at the end of the phase.

Fig. 6 shows the phase–amplitude (PA) response for PRPDs in 2D representation, where amplitudes of 3600 power cycles are accumulated to generate the PA response, and the number of PDs per 3600 power cycles is illustrated by different colors in a 2D representation. The PA response for PRPDs is defined as

$$\mathbf{X}_{in} = \begin{pmatrix} x_{in}(1,1) & x_{in}(1,2) & \cdots & x_{in}(1,P) \\ x_{in}(2,1) & x_{in}(2,2) & \cdots & x_{in}(2,P) \\ \vdots & \vdots & \ddots & \vdots \\ x_{in}(K,1) & x_{in}(K,2) & \cdots & x_{in}(K,P) \end{pmatrix}, \quad (2)$$

where $K = 256$, $\sum_{k=1}^{K} x_{in}(k,p) = M$, and $x_{in}(k,p) \in \{0, 1, \cdots, M\}$ is the number of cycles with an amplitude value k at the pth phase with $k = 1, 2, \ldots, K$ and $p = 1, 2, \ldots, P$. Sequential PRPDs in (1) are transformed into the PA response in (2), so the size of the data is substantially decreased. It can be clearly seen from Fig. 6a and Fig. 6f that the PD activity of the defects protruding from the electrode and void are almost symmetric in two half-circles. In addition, the distribution of the fault floating activity, which is not symmetric, is also presented in Fig. 6b.
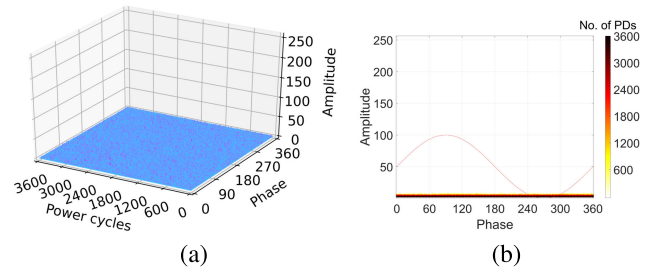
C. NOISE MEASUREMENTS

The noise was measured in a laboratory environment using the UHF drain valve sensor. Fig. 7 presents an example of noise signals for 3600 power cycles and the corresponding PA response in a 2D representation. Here, it can be seen that with small amplitudes, the pulses of noise signals are sparse and scattered without regularity in all ranges of phases and power cycles, as shown in Fig. 7.

III. PROPOSED CNN FOR PD DIAGNOSIS

In this section, we focus on CNN-based PRPD fault diagnosis. The proposed CNN architecture for classifying PRPDs in power transformers is shown in Fig. 8. The structure of the model is comprised of an input layer, convolutional layers, max-pooling, dropout, batch normalization, flatten, fully connected layers, and an output layer. In the input layer, the PA response, $\mathbf{X}_{in}$, in (2) is used for PRPDs, where the matrix size of $\mathbf{X}_{in}$ is smaller than that of sequential data $\mathbf{X}$ in (1).

Table 3 shows the detailed structure of the proposed CNN model, where the total parameters for the training set number 145,331. To acquire the appropriate characteristics, the input
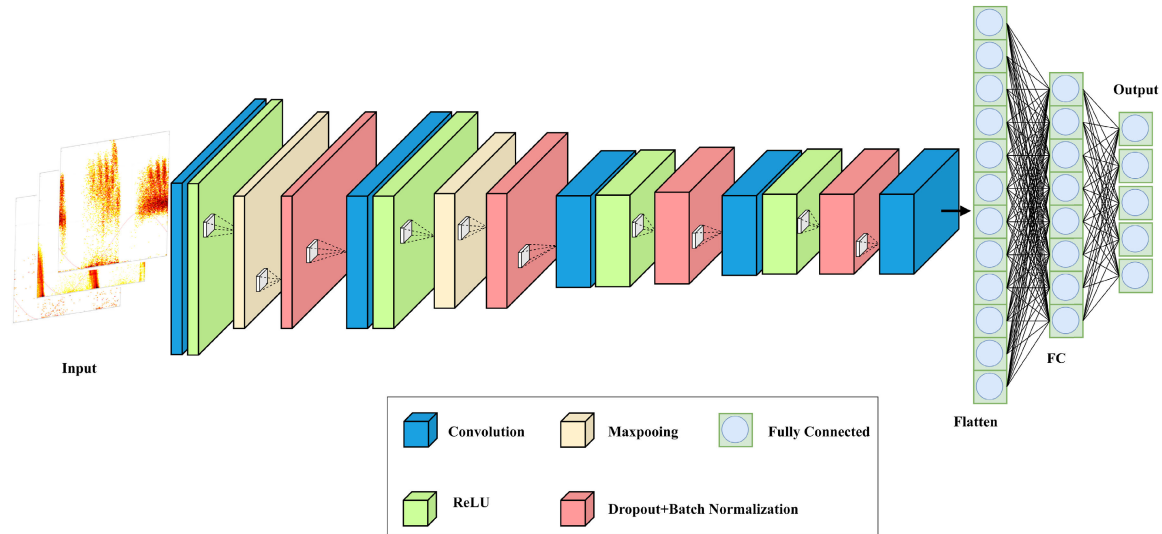
FIGURE 8. Proposed CNN architecture.

data will be resized according to the designed CNN model. The convolutional layer is utilized directly to extract features from the input and map extracted features to form new feature maps. Specifically, using a convolution operation, the convolutional layers convolve the input of the local regions to extract the local features, which is usually referred to as weight-sharing and is afterward followed by the activation function to produce features of the output. It can be seen from Table 3 that, in all convolutional layers, we employed multiple filters with a kernel size of 5 × 5 and 3 × 3, which have the function of reducing the input size for the next convolution layer resulting in less computational complexity for the next network layer. In addition, it also plays a leading role in reducing the number of parameters, and consequently, the training efficiency of the model is substantially improved. The activation function, which permits the network to obtain a nonlinear expression of the input to intensify the representation capacity and make the learned features increasingly dividable, is utilized. Owing to both the simplicity of the implementation and its good performance on a variety of predictive tasks, the rectified linear unit (ReLU), which accelerates the convergence of the CNNs, has become the most popular choice of activation function in recent years [32], [33]. The ReLU is defined as follows

$$f(z^l) = \max\{0, z^l\} = \begin{cases} z^l, & z^l \geq 0, \\ 0, & \text{else,} \end{cases} \qquad (3)$$

in which z^l is an element of the outputs in the l-th convolutional layer. After the first two convolutional layers, the max-pooling layer, which plays the key role of increasing invariance to small local translations as well as decreasing the spatial size of the features and the number of subsequent learnable parameters of the network by subsampling, is employed [33], [34]. The max-pooling acts as a down-sampling function with a view to reducing the size of the feature layout and obtaining location-invariant features. As

can be seen in Table 3, with a 2 × 2-size window and the sampling window step size of 2, pooling layers separate the previous layer's output into a nonoverlapping subregion and conducts a local max operation over the input features. With the proposed CNN model, after the convolution, ReLU, and pooling layer, dropout, and batch normalization which are used as regularization techniques to reduce the over-fitting and accelerate the training process of the network [35], [36]. By permitting some fraction of the nodes in each layer to be literally rejected before calculating the subsequent layer in each iteration during the training, the dropout technique handles the challenge of a large number of parameters, leading to an acceleration of the training process and the prevention of an overmatching. Likewise, batch normalization, a popular and effective technique that consistently accelerates the convergence of the networks, was also employed to avoid vanishing and exploding gradients mainly caused by the internal covariance shift throughout the training.

Finally, the classification layer, commonly called a softmax layer, comes after the fully connected layer. In this layer, the output value was designed to be equal to the number of objects to be classified, as can be seen in Table 3. The value was set to 7. The task here is to determine the probability of the input batch elements belonging to one of the categories and give the final probabilities for each label. During the classification process, an output value interval is produced with a range of 0–1 for seven different objects, which means that the output to be interpreted as the probability that a given item belongs to a particular class, based on the largest output value of the class, can then be chosen. This means that the output that produces a value of near 1 is perceived to be the object predicted by the network. At the output layer, the output for C classes is obtained using a softmax activation function as

$$\mathbf{z} = [z_1, \cdots, z_C]^T = \sigma(\mathbf{h}), \qquad (4)$$

TABLE 4. Experimental dataset for PRPDs and noise.

Fault type	(0)	(1)	(2)	(3)	(4)	(5)	(6)
Number of experiments	394	460	366	381	377	379	379

where z_j is the predicted interference representing the j-th category in the C classes, $\mathbf{h} = [h_1, \cdots, h_C]^T$ is the output of the last fully connected layer, and $\sigma(\mathbf{h})$ is the softmax function, which is defined as

$$z_j = [\sigma(\mathbf{h})]_j = \frac{e^{h_j}}{\sum_{i=1}^{C} e^{h_i}}. \qquad (5)$$

During training, the parameters of the proposed CNN were learned through the minibatch $\mathcal{B}$ to minimize the following loss function

$$J(\Theta) = \frac{1}{|\mathcal{B}|} \sum_{b \in \mathcal{B}} Loss(b), \qquad (6)$$

where Θ is denoted as a correction of every learnable parameter in the model, and $|\cdot|$ is the number of elements in a set. In (5), the loss function is calculated based on the cross-entropy loss as follows

$$Loss(b) = -\sum_{i=1}^{C} t_i \log\left(z_i^{(b)}\right), \qquad (7)$$

where the superscript (b) is the index for the b-th training sample in the minibatch $\mathcal{B}$, $t_i = 1$ when the index i is the index for the ground truth, and $t_i = 0$ otherwise. To minimize the loss function, stochastic gradient decent optimization algorithms such as AdaGrad, AdaDelta, and Adam [37]–[39] are used. Here, the Adam optimizer is used to update the network-learnable parameters.

IV. PERFORMANCE EVALUATIONS

In this section, the performance evaluation of partial discharge pattern recognition using PRPDs and noise measurements is demonstrated to clarify the results of the CNN algorithm for fault diagnosis in power transformers. The number of samples for noise and each defect in the experiments is shown in Table 4, where noise and six types of faults (protruding electrode, floating, particle, surface, turn-to-turn, and void) are considered and numbered as 0, 1, 2, 3, 4, 5, and 6, respectively.

In our experiments, all of the data was typically split into three parts (training set, validation, and test set) with 80%, 10%, and 10% samples of the data, respectively. Therefore, with 2736 samples in the entire dataset, the training database possesses 2188 samples in total. The validation and test set take up the same figure, at 274 each. During the training process, the optimization step was accomplished to acquire the optimized hyperparameters in accordance with the batch size, layer type, number of layers, and filter size. The proposed model was trained by utilizing different combinations

TABLE 5. Maximum and minimum bound of hyperparameter optimization.

Hyperparameter	Minimum	Maximum	Type
Minibatch size	8	128	Integer
Number of layers	6	18	Integer
Kernel size	3×3	7×7	Integer
Drop factor	0.3	0.5	Real
Learning rate	0.0001	0.005	Real

TABLE 6. Performance evaluation with different representations of data.

Data type	Sequential data (%)	PA response (%)
Protruding electrode	97.50	100
Floating	97.83	100
Particle	100	100
Surface	100	100
Turn to turn	100	97.37
Void	92.11	100
Noise	100	100
Overall	98.18	99.64

of parameters, and the best combination was chosen. We carried out trials to alleviate the effects of random initial values on the network during our experiments, and then averaged the results to confirm the robustness of the proposed model. Table 5 shows the optimization of the maximum and minimum boundary range of each hyperparameter and whether the parameter was an integer or a real value. Among all models generated, the 16-layer CNN with a 5×5-size kernel in the first convolutional layer and a 3×3-size kernel in the rest of the layers using a learning rate, an exponential decay rate for the first moment, a minibatch size, and a learning rate drop factor of 0.0002, 0.5, 32, and 0.5, respectively, achieves the highest overall PD pattern recognition accuracy and the highest pattern recognition accuracy for every type of defect. In addition, the types of activation functions applied strongly affect the PD pattern recognition performance of the CNN. In this study, the performance of CNN was determined with different activation functions such as Tanh, Sigmoid, Swish, ReLU, and Leaky ReLU using the same optimized hyperparameters. The results show that the ReLU function has a higher pattern recognition accuracy than those of CNNs with the other four activation functions.

Table 6 presents the classification results of the performance evaluation of the proposed model based on different representations of the dataset. It can be seen that the proposed CNN model with both types of input data, sequential data, and PA response exhibit the appreciate results for the classification of the transformers' discharge faults. Nevertheless, the proposed CNN when using the PA response achieves a 99.64% overall accuracy, which is 1.46% higher than when using the sequential data. Here, we use the PA response in (2) as the input of the proposed CNN. Regarding each defect,
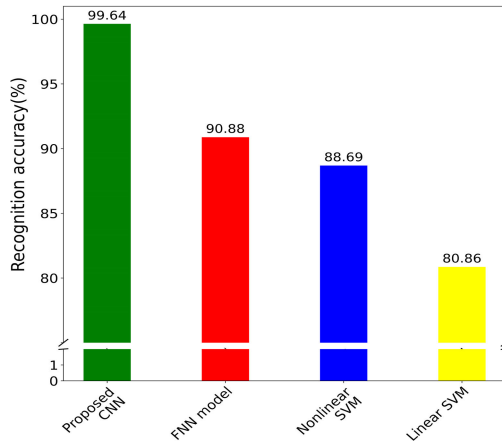
FIGURE 9. Classification accuracy performance for different models.

from the table, it is obvious that the PD pattern recognition accuracy rate of the proposed CNN method utilizing the PA response as the input shows better results for most of the transformer incipient faults. However, as can be seen, with the "turn-to-turn" defect, the proposed model utilizing continuous data exhibits better characteristics than that using PA data. An analysis of the results shows that the proposed model also has the advantage of using continuous data in this case.

Fig. 9 provides a comparison of the evaluation performance of the proposed CNN model, FNN, and SVMs for partial discharge pattern recognition in a power transformer, where the SVMs use maximum values for power cycles at each phase as a feature vector. For comparison, we used linear and nonlinear SVMs with a radial basis function (RBF) and an FNN model as the baseline models [40], [41]. In the SVMs, the feature vector was obtained based on the maximum of the amplitudes from the PRPD, where the parameter $C = 0.01$ for the linear SVM and the parameters $C = 0.01$ and $\gamma = 0.1$ were chosen after parameter estimation using a grid search. The FNN model was designed with seven hidden layers, the total parameters of which were 4,203,847. In comparison with the proposed CNN method, it is clear that the FNN requires more learnable parameters despite the smaller number of hidden units than the CNN model. From Fig. 9, it is clearly observed that the proposed CNN model attained the highest overall classification accuracy performance (at roughly 99.64%), followed by the FNN and the linear and nonlinear SVM models, of which the linear SVM model is the least effective among the approaches. Note that the FNN is superior to the SVMs, and the nonlinear SVM with RBF is somewhat superior to the linear SVM. This is because the improvement in performance comes from the ability to automatically acquire characteristics of the PA response of PRPDs and recombine them to generate new features from the raw input. Meanwhile, the FNN used the raw input without the phase information of the PRPDs, and the SVMs used a manually created feature vector that combined the characteristics of PRPDs.
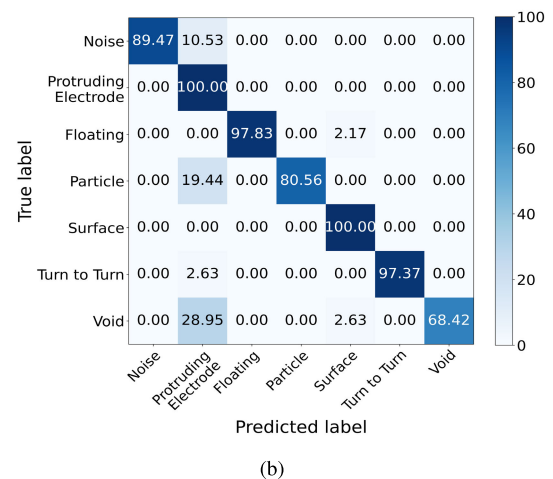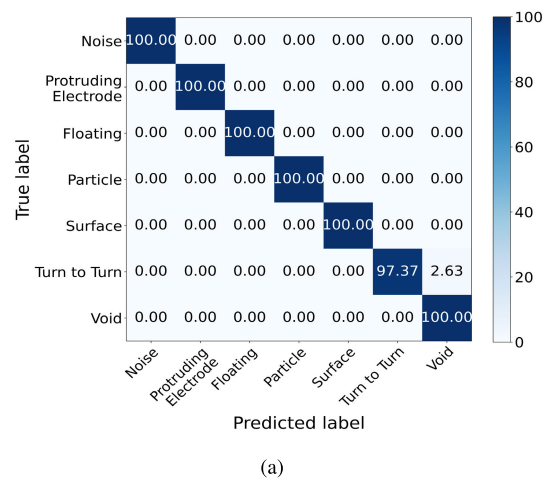

(a)

(b)

FIGURE 10. Confusion matrix for (a) proposed CNN and (b) FNN models.

Using a PA response, the PD pattern recognition for the FNN and SVMs was also investigated, as shown in Table 7. It is clear that the accuracy rate of the PD pattern recognition of the proposed CNN for every type of transformer incipient fault is the highest among all methods, with an accuracy rate of 100% for almost all defect types. Meanwhile, with only 80.56% and 68.42% accuracy rates in recognizing the particle and void defects, respectively, the FNN model proved less viable in discriminating the defect types. Likewise, for the SVM, the linear and nonlinear versions are both capable of differentiating several faults, such as a protruding electrode and a surface defect with a nonlinear SVM, and the turn-to-turn defect and noise with a linear model. In particular, the protruding electrode and void classification accuracies of 40% and 57.89% are lower than the 60% and 42.11% of the proposed CNN, respectively, which indicates the inefficiency of the linear SVM model in the transformer fault classification.

Furthermore, in order to assert the advantages of the proposed CNN technique and to more intuitively reflect the accuracy of the classification results, the confusion matrix of the FNN is presented in Fig. 10. It can be seen in Fig. 10 that the proposed model shows the degree of distinction of
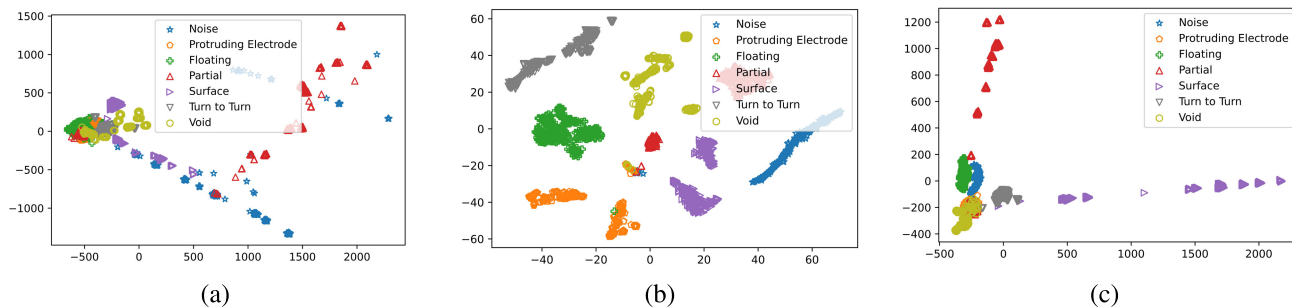
FIGURE 11. Data visualization using t-distributed stochastic neighbor embedding (t-SNE) of 2188 training samples for (a) input data, (b) FC layer of proposed CNN, and (c) FC layer of FNN model.

TABLE 7. Comparison of classification performance with other methods using PA response.

Fault types	Linear SVM (%)	Nonlinear SVM with RBF (%)	FNN model (%)	Proposed CNN (%)
Protruding Electrode	40.0	100	100	100
Floating	86.96	93.48	97.83	100
Particle	91.67	80.56	80.56	100
Surface	92.11	100	100	100
Turn to turn	97.37	86.84	97.37	97.37
Void	57.89	68.42	68.42	100
Noise	100	89.47	89.47	100
Overall	80.66	88.69	90.88	99.64

defect types as relatively high, with accuracy rates of 100% for the defect types of protruding electrode, floating, particle, surface, and void, and approximately 97.4% for the turn-to-turn fault, making it easier to differentiate the defect types. In addition, from Fig. 10a, we can see that among all of the faults, only the turn-to-turn defect is misidentified and incorrectly recognized as a void defect with 2.63%, indicating that the characteristics of the proposed model extraction perform well in other faults but lack the ability to identify the defect in turn-to-turn. This is because there is one sample of 38 samples in the test set according to the turn-to-turn defect, which has the same distribution as the void defect. Meanwhile, the CNN model has shown undisputed efficiency in distinguishing most of the faults. The FNN method has been perceived as less effective in identifying the defect types of protruding electrode, void, and noise. This performance is lower than that of the CNN model at 15%, 5.26%, and 10.53%, respectively.

To better understand the CNN model, we analyzed the internal representation of the trained network at the fully connected layer. Figs. 11a, 11b, and 11c show the t-distributed stochastic neighbor embedding (t-SNE) representations of vectors for the input and fully connected layers of the proposed CNN and FNN schemes, respectively. The t-SNE embedded high-dimensional vectors into 2D spaces while retaining the pairwise similarity [42]. As can be seen in Fig. 11a, the PD signals are very close to each other, so it is
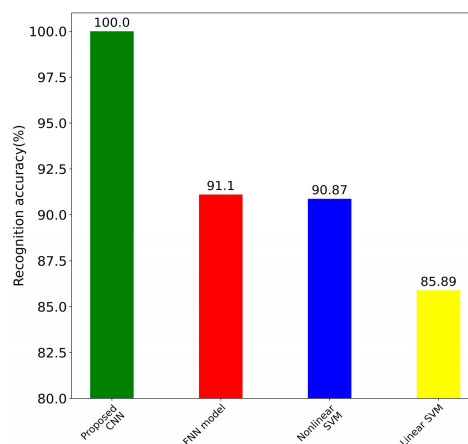

FIGURE 12. Classification accuracy performance for different models in case of without noisy data.

TABLE 8. Training and testing time comparisons.

Model type		Training time (min)	Test time on test set (s)
Using PA response	Linear SVM	0.034	∼ 0.009
	Nonlinear SVM	0.0164	∼ 0.110
	FNN model	0.89	∼ 0.111
	Proposed CNN	2.93	∼ 0.161
Proposed CNN using PRPD		35.93	∼ 1.05

difficult to identify and recognize discharge defects precisely using the PD input signal. By contrast, Fig. 11b and Fig. 11c show that the vector of the fully connected layer of both models was much more dispersed when compared to the input vector. The vectors of the fully connected layer of the proposed CNN model for some data of the protruding electrode, floating, particle, surface, turn-to-turn, and void faults in the power transformer are similar to those for some noise data. In addition, the clear distinct distribution of the incipient faults of the transformers is illustrated in Fig. 11b,
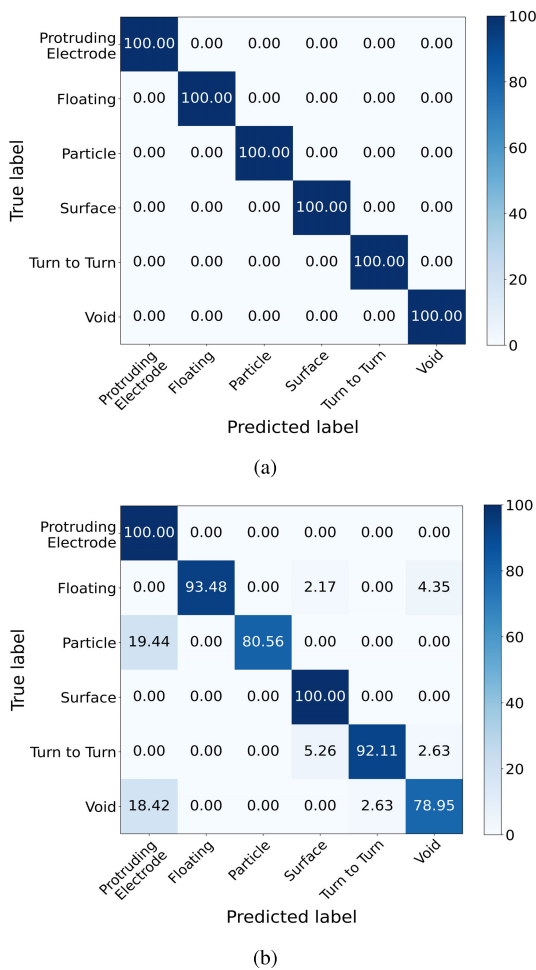
FIGURE 13. Confusion matrix for (a) proposed CNN and (b) FNN models in case of without noisy data.

which indicates the predominance of the proposed CNN over the FNN model. Therefore, our proposed method can serve as an effective partial discharge pattern recognition method for a power transformer.

Table 8 shows the training and testing time comparisons for the FNN, SVMs, and proposed CNN, where the timing was normalized to a hypothetical 1-GHz single-core CPU to make the measurement meaningful. In our experiments, with the same hardware configuration, the models were trained and tested on an NVIDIA Titan X GPU with 3584 cores, each running at 1.4 GHz. As can be seen, the proposed CNN model was slower with a 2.93-min training time compared to 0.89 and 0.034 min of the FNN and linear SVM model, respectively. However, a test period of roughly 0.161 s for the proposed model using a PA response is appropriate for an offline application. Moreover, to employ a different input, the use of a PA response as an input of the proposed CNN outperforms that of a PRPD with less than 33 min required for the training and 0.889 s required for the testing. This leads to the advantage of the proposed model using a PA response, not only improving the performance of the pattern recognition but also appropriately reducing the size of the input leading

to a reduction of the complexity of the model and the memory required for its deployment.

To enhance the adaptability to real-world situations, the pattern recognitions of an FNN, SVMs, and the proposed PD classified model were also investigated by comparing different partial discharge models without noise. The overall classification results of different models using partial discharge data without noise are illustrated in Fig. 12. Although the proposed model exhibits the highest overall PD pattern recognition accuracy of 100%, followed by the FNN and non-linear SVMs, which have a similar capability of classifying the partial discharge defects without noise with accuracy rates of 91.1% and 90.87%, respectively, the linear SVM model is the lowest among the different approaches. In addition, Fig. 13 shows the PD pattern recognition accuracy for each type of PD fault without noise using the confusion matrix. As shown in Fig. 13a, the proposed CNN model can successfully predict PD patterns for every type of defect without a noise signal with an accuracy of 100%. Meanwhile, the FNN model shows a lower efficiency with only two detected faults and is inefficient in predicting void and particle defects with accuracy rates of 78.95% and 80.56%, as shown in Fig. 13b. A comparison of the different methods, including the FNN and SVMs, under different cases of noise and no noise, indicates that the proposed approach is reasonable and achieves an improvement over the existing methods used for PD classification under both noisy and quiet conditions, which implies that the proposed model exhibits the strongest tolerance against noise contamination.

V. CONCLUSION

In this article, we proposed a CNN-based fault diagnosis method to detect defects in power transformers. PRPDs were obtained by utilizing a new UHF drain sensor and included six types of faults: protruding electrode, particle, floating, surface, bad contact between windings, and void. The proposed model uses the PA response from PRPDs for the dimension reduction of the CNN input. The experimental results revealed that the proposed CNN achieved a classification accuracy of 99.64% and had 18.78%, 10.95%, 8.76%, and 1.46% higher classification performance than linear SVM, nonlinear SVM, FNN, and CNN using PRPDs, respectively. The proposed CNN using the PA response can be applied to the fault diagnosis of other power equipment. In addition, further verification of the proposed method will be conducted using continuous data and the PA response based on mixed simulation PD signals and the actual transformer PD signals when considering real-world situations.

REFERENCES

[1] C. Wang, J. Wu, J. Wang, and W. Zhao, "Reliability analysis and overload capability assessment of oil-immersed power transformers," *Energies*, vol. 9, no. 1, p. 43, Jan. 2016.

[2] A. Jahromi, R. Piercy, S. Cress, J. Service, and W. Fan, "An approach to power transformer asset management using health index," *IEEE Elect. Insul. Mag.*, vol. 25, no. 2, pp. 20–34, Mar. 2009, doi: 10.1109/MEI.2009.4802595.

[3] E. Gulski, "Digital analysis of partial discharges," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 2, no. 5, pp. 822–837, Oct. 1995, doi: 10.1109/94.469977.

[4] T. Wen, Q. Zhang, J. Ma, X. Liu, Z. Wu, L. Zhang, J. Zhao, N. Shimomura, and W. Chen, "Research on the detecting effectiveness of on-site lightning impulse test for GIS equipment with insulation defects," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 25, no. 2, pp. 551–558, Apr. 2018, doi: 10.1109/TDEI.2018.006961.

[5] Y.-B. Wang, D.-G. Chang, Y.-H. Fan, G.-J. Zhang, J.-Y. Zhan, X.-J. Shao, and W.-L. He, "Acoustic localization of partial discharge sources in power transformers using a particle-swarm-optimization-route-searching algorithm," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 24, no. 6, pp. 3647–3656, Dec. 2017, doi: 10.1109/TDEI.2017.006857.

[6] M. Mondal, G. B. Kumbhar, and S. V. Kulkarni, "Localization of partial discharges inside a transformer winding using a ladder network constructed from terminal measurements," *IEEE Trans. Power Del.*, vol. 33, no. 3, pp. 1035–1043, Jun. 2018, doi: 10.1109/TPWRD.2017.2683560.

[7] C. Pan, J. Tang, Y. Zhang, X. Luo, and X. Li, "Variation of discharge characteristics with temperature in moving transformer oil contaminated by metallic particles," *IEEE Access*, vol. 6, pp. 40050–40058, 2018, doi: 10.1109/ACCESS.2018.2856580.

[8] J. Singh, Y. Sood, and R. Jarial, "Condition monitoring of power Transformers–Bibliography survey," *IEEE Elect. Insul. Mag.*, vol. 24, no. 3, pp. 11–25, May 2008, doi: 10.1109/MEI.2008.4591431.

[9] S. Chakravorti, D. Dey, and B. Chatterjee, *Recent Trends in the Condition Monitoring of Transformers: Theory, Implementation, and Analysis*. Cham, Switzerland: Springer, 2013.

[10] H. Dong, X. Yang, A. Li, Z. Xie, and Y. Zuo, "Bio-inspired PHM model for diagnostics of faults in power transformers using dissolved Gas-in-Oil data," *Sensors*, vol. 19, no. 4, p. 845, Feb. 2019.

[11] *High-Voltage Test Techniques-Partial Discharge Measurements*, International Electrotechnical Commission, Geneva, Switzerland, 2000.

[12] M. D. Judd, L. Yang, and I. B. B. Hunter, "Partial discharge monitoring of power transformers using UHF sensors—Part I: Sensors and signal interpretation," *IEEE Elect. Insul. Mag.*, vol. 21, no. 2, pp. 5–14, Mar. 2005, doi: 10.1109/MEI.2005.1412214.

[13] M. D. Judd, L. Yang, and I. B. B. Hunter, "Partial discharge monitoring for power transformer using UHF sensors—Part 2: Field experience," *IEEE Elect. Insul. Mag.*, vol. 21, no. 3, pp. 5–13, May 2005, doi: 10.1109/MEI.2005.1437603.

[14] K. Raja, F. Devaux, and S. Lelaidier, "Recognition of discharge sources using UHF PD signatures," *IEEE Elect. Insul. Mag.*, vol. 18, no. 5, pp. 8–14, Sep. 2002, doi: 10.1109/MEI.2002.1044316.

[15] L. Yang and M. D. Judd, "Recognising multiple partial discharge sources in power transformers by wavelet analysis of UHF signals," *IEE Proc. Sci., Meas. Technol.*, vol. 150, no. 3, pp. 119–127, May 2003, doi: 10.1049/ip-smt:20030201.

[16] D. Aschenbrenner, H.-G. Kranz, W. R. Rutgers, and P. van den Aardweg, "On line pd measurements and diagnosis on power transformers," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 12, no. 2, pp. 216–222, Apr. 2005, doi: 10.1109/TDEI.2005.1430392.

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

[18] J. Faiz and M. Soleimani, "Assessment of computational intelligence and conventional dissolved gas analysis methods for transformer fault diagnosis," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 25, no. 5, pp. 1798–1806, Oct. 2018.

[19] Z. Chen, T. Lin, N. Tang, and X. Xia, "A parallel genetic algorithm based feature selection and parameter optimization for support vector machine," *Sci. Program.*, vol. 2016, no. 2, pp. 1–10, Jun. 2016.

[20] Y. Zhang, H. Wei, R. Liao, Y. Wang, L. Yang, and C. Yan, "A new support vector machine model based on improved imperialist competitive algorithm for fault diagnosis of oil-immersed transformers," *J. Electr. Eng. Technol.*, vol. 12, no. 2, pp. 830–839, Mar. 2017.

[21] Y. Zhang, J. Liu, H. Zheng, and K. Wang, "Feasibility of a universal approach for temperature correction in frequency domain spectroscopy of transformer insulation," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 25, no. 5, pp. 1766–1773, Oct. 2018.

[22] J. L. Guardado, J. L. Naredo, P. Moreno, and C. R. Fuerte, "A comparative study of neural network efficiency in power transformers diagnosis using dissolved gas analysis," *IEEE Trans. Power Del.*, vol. 16, no. 4, pp. 643–647, Oct. 2001.

[23] V. Miranda and A. R. G. Castro, "Improving the IEC table for transformer failure diagnosis with knowledge extraction from neural networks," *IEEE Trans. Power Del.*, vol. 20, no. 4, pp. 2509–2516, Oct. 2005.

[24] D. Cireäan, U. Meier, J. Masci, and J. Schmidhuber, "Multi-column deep neural network for traffic sign classification," *Neural Netw.*, vol. 32, pp. 333–338, Aug. 2012.

[25] X. Liu, Z. Deng, and Y. Yang, "Recent progress in semantic image segmentation," *Artif. Intell. Rev.*, vol. 52, no. 2, pp. 1089–1106, Aug. 2019.

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

[27] A. Ucar, Y. Demir, and C. Guzelis, "Object recognition and detection with deep learning for autonomous driving applications," *Simulation*, vol. 1, pp. 1–11, Dec. 2017.

[28] H. Mirzaei, A. Akbari, E. Gockenbach, and K. Miralikhani, "Advancing new techniques for UHF PDdetection and localization in the power transformers in the factory tests," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 22, no. 1, pp. 448–455, Feb. 2015, doi: 10.1109/TDEI.2014.004249.

[29] M. Siegel, M. Beltle, S. Tenbohlen, and S. Coenen, "Application of UHF sensors for PD measurement at power transformers," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 24, no. 1, pp. 331–339, Feb. 2017, doi: 10.1109/TDEI.2016.005913.

[30] J. Vedral and M. Kriz, "Signal processing in partial discharge measurement," *Metrology Meas. Syst.*, vol. 17, no. 1, pp. 55–64, Jan. 2010.

[31] G.-J. Zhang, Y.-B. Wang, D.-G. Chang, X.-J. Shao, J.-Y. Zhan, and W.-L. He, "Separation of multiple partial discharge sources in power transformer," in *Proc. Int. Conf. High Voltage Eng. Power Syst. (ICHVEPS)*, Oct. 2017, pp. 12–17, doi: 10.1109/ICHVEPS.2017.8225857.

[32] A. Fred Agarap, "Deep learning using rectified linear units (ReLU)," 2018, *arXiv:1803.08375*. [Online]. Available: http://arxiv.org/abs/1803.08375

[33] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Neural Inf. Process. Syst.*, 2012, pp. 84–90.

[34] D. Ciresan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3642–3649, doi: 10.1109/CVPR.2012.6248110.

[35] C. Garbin, X. Zhu, and O. Marques, "Dropout vs. Batch normalization: An empirical study of their impact to deep learning," *Multimedia Tools Appl.*, vol. 79, nos. 19–20, pp. 12777–12815, May 2020.

[36] M. Z. Alom, T. M. Taha, C. Yakopcic, S. Westberg, P. Sidike, M. S. Nasrin, M. Hasan, B. C. Van Essen, A. A. S. Awwal, and V. K. Asari, "A state-of-the-art survey on deep learning theory and architectures," *Electronics*, vol. 8, no. 3, p. 292, Mar. 2019.

[37] J. Duchi, E. Hazan, and Y. Singer, "Adaptive subgradient methods for online learning and stochastic optimization," *J. Mach. Learn. Res.*, vol. 12, pp. 2121–2159, Feb. 2011.

[38] M. D. Zeiler, "ADADELTA: An adaptive learning rate method," 2012, *arXiv:1212.5701*. [Online]. Available: http://arxiv.org/abs/1212.5701

[39] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–41.

[40] C. Cortes and V. N. Vapnik, "Support vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1997.

[41] H.-Y. Huang and C.-J. Lin, "Linear and kernel classification: When to use which?" in *Proc. SIAM Int. Conf. Data Mining*, Jun. 2016, pp. 216–224.

[42] L. van der Maaten and G. Hinton, "Visualizing high-dimensional data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

THE-DUONG DO received the B.E. degree in mechatronics from the Hanoi University of Science and Technology, Hanoi, Vietnam, in 2018. He is currently pursuing the master's degree with the Information Technology Convergence Laboratory, Department of Electronic Engineering, Myongji University, South Korea. His major research interests include deep learning applications for fault diagnosis, generative adversarial nets, and automotive radar signal processing.

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

YONG-SUNG CHO received the B.S. and M.S. degrees in electrical engineering from Busan National University, Busan, South Korea, in 2003 and 2006, respectively. Since 2006, he has been with the Power Apparatus Research Center, Korea Electrotechnology Research Institute, where he is currently a Senior Researcher. His research interests include multi-physics simulation and partial discharge diagnostics system for power apparatus.

JONG-HO SUN was born in South Korea in 1962. He received the B.S. and Ph.D. degrees in electrical engineering from Pusan National University in 1986 and 2001, respectively. Since 2001, he has been with the Korea Electrotechnology Research Institute, where he is currently a Principal Researcher. His interests are the diagnosis and monitoring technologies of electrical power equipment.

YONG-HWA KIM (Member, IEEE) received the B.S. degree in electrical engineering and the Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 2001 and 2007, respectively. From 2007 to 2011, he was a Senior Researcher with the Korea Electrotechnology Research Institute, Geonggi, South Korea. From 2011 to 2013, he was an Assistant Professor with the Division of Maritime Electronic and Communication Engineering, Mokpo National Maritime University, South Korea. In 2013, he joined the Faculty with the Department of Electronic Engineering, Myongji University, South Korea. His general research interests include communication systems, fault diagnosis, and digital signal processing, with an emphasis on artificial intelligence for communications, radar systems, and smart grids.

. . .