Anomaly Detection for Partial Discharge in Gas-Insulated Switchgears Using Autoencoder

NGOC-DIEM TRAN THI¹, THE-DUONG DO¹, JAE-RYONG JUNG², HYANGEUN JO², AND YONG-HWA KIM¹, (Member, IEEE)
¹Department of Electronic Engineering, Myongji University, Yongin 17058, South Korea
²Research and Development Center, Hyosung Corporation, Changwon 04144, South Korea
Corresponding author: Yong-Hwa Kim (yongkim@mju.ac.kr)

This work was supported in part by the Korea Electric Power Corporation under Grant R18XA01, and in part by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) and the Ministry of Trade, Industry, and Energy (MOTIE) of the Republic of Korea under Grant 20179310100050.

ABSTRACT In this article, we propose a new anomaly detection method to detect the partial discharge in a gas-insulated switchgear. An autoencoder was used for anomaly detection and was modeled on the one-class classification problem. Based on the one-class classification scenario, in which the training data exploited the noise data only, the proposed autoencoder learned the low-dimensional latent information from the high-dimensional space of the input signal. Then, the reconstruction error was used as a fault indicator, and the threshold was determined using the partial discharge data. The performance of the proposed AE was verified by on-site noise and PRPD experiments, using an online UHF PD monitoring system in the real-world environment. The results showed that the proposed autoencoder not only achieved 86.75% detection performance for the on-site noise and partial discharge data in gas-insulated switchgears but also allowed better detection performance than the one-class support vector machine learning procedure by 40.5%.

INDEX TERMS Partial discharge (PD), fault detection, gas-insulated switchgear (GIS), anomaly detection, autoencoder (AE).

I. INTRODUCTION
Power systems are rapidly expanding with increasing power demands and distributed energy resources. For utilities, owing to the aging of existing power systems, asset management is a requisite to extend the life of infrastructure assets and ensure the reliability of the power grid [1]. Several studies, such as event detection [2], reliability evaluation [3], [4], and fault diagnosis [5] have been conducted for asset management in power grids. Condition monitoring and diagnosis is a major part of asset management. Thus, online and offline measurements are performed to monitor the conditions of the power grid assets [5], [6]. A gas-insulated switchgear (GIS), applied to substations, is a major protection device for electric power facilities. A GIS is a valuable device in protecting, controlling, and isolating the equipment in a power grid in the case of an incident (e.g., power surge) [7], [8]. Internal defects can occur in a GIS during the process of transfer, installation, and operation [7]. When a failure occurs in a GIS, the impact of the accident is huge so recovery takes a lot of time and the power outage also increases. Defects in GISs cause PDs that result in the breakdown of insulation [8]. Therefore, it is essential to avoid failure by detecting the PD of a GIS and addressing the defects in the GIS at an early stage [9], [10].

According to IEC 62478, there are several phenomena, such as light emissions, acoustic waves, electromagnetic signals, and chemical reactions caused by a PD occurrence in a GIS [11]. To measure these phenomena, various electrical, mechanical, and chemical methods have been used. Few existing electrical methods use ultra-high frequency (UHF) sensors, the acoustic methods use the acoustic sensors, and the chemical methods use the dissolved gas analysis technique [10], [12], [13]. In particular, the electrical method that uses the UHF sensors has the advantage of high sensitivity for PD detection. In this study, the UHF method was utilized in the PD measurement system for condition monitoring and assessment of GISs [12].

In order to investigate the PD characteristics in a GIS, the time-resolved partial discharge (TRPD) and the phase-resolved partial discharge (PRPD) analysis methods were
studied [14]. The TRPD based method analyzed the time-domain, frequency-domain, and time-frequency-domain features from the PD pulses [14]–[16]. The PRPD based method measured and analyzed the amplitude at each phase. Defect types were identified by analyzing the number of the PD pulses, maximum amplitude, or average amplitude in each phase [17]. In the PRPD method, we used the signal-processing techniques in the time-domain [18], frequency-domain [19], and time-frequency-domain [20] to obtain meaningful features. From these features, machine learning based classifiers, such as neural networks [21], decision trees [22], k-nearest neighbors (k-NN) [23], and support vector machines (SVMs) [24] were trained for PD classifications.

To improve the performance of fault detection, many deep learning models have been studied to extract features and classify the PDs automatically in an end-to-end manner. Deep neural networks have achieved state-of-the-art accuracies in multiple pattern recognition tasks in different domains, such as computer vision, speech recognition and text classification [25]–[27]. To improve the fault detection accuracy using the PRPD method, various deep neural network models have been proposed based on popular neural network structures such as the convolutional neural network (CNN) [28], long short-term memory (LSTM) [29], and self-attention network [30]. However, most of the existing deep learning-based fault diagnosis methods are supervised, i.e., they require training the data with their corresponding fault labels [31], [32]. It is generally difficult to obtain an on-site labeled fault data.

In this article, to overcome the lack of on-site labeled fault data for supervised learning, we propose a new anomaly detection method for GISs using autoencoders (AEs). An AE is a simple feed-forward neural network with multiple layers of hidden nodes, where the number of hidden nodes is usually fewer than the input nodes. The model trains the hidden nodes to reconstruct the input at the output without a labeled dataset by minimizing the reconstruction error [33]. The proposed AE learnt the features directly from the unlabeled noise data during the training process. Noise and PRPD measurements of GISs were conducted on-site. Noise data was used for training, validating, and testing as well as in new test sets. The loss function was used to calculate the reconstruction errors in our AE model and the hyper parameters of the proposed AE were determined using a validation set. The test set, which was composed of noise data and PD data, was used to determine the threshold for fault detection in a semi-supervised manner. Then, the detection performance was verified using a new test set that included the on-site noise and the PRPD data. The major contributions of this article are summarized as follows:

- Anomaly detection using AEs was applied for the first time to detect faults in semi-supervised learning. The proposed AE was trained using noise data in the real-world environments and fault data was used to determine the threshold. The proposed AE had an advantage in detecting faults in real-world applications because the proposed AE used only the noise data during the training process.
- The proposed AE outperformed the one-class support vector machine (OCSVM), an anomaly detection method that required only one class of normal samples [34]. This was because AEs had the advantage of feature extraction based on the on-site noise data for GISs.
- The performance of the proposed AE was verified through on-site noise and PRPD experiments in the real-world environments. On-site PRPD data included seven types of faults that could occur in a GIS, such as the crack, floating, free particle, protrusion on conductor (POC), protrusion on enclosure (POE), particle on spacer (POS), and void. By using the on-site noise and the PRPD data, the proposed method exhibited an improved detection performance than the OCSVM by 40.5% and also exhibited a detection performance of 86.75% for on-site noise and PD data in GISs.

The remainder of this article is organized as follows. We briefly introduce the anomaly detection in Section II. Section III presents the on-site noise and the PD measurements for GISs. In Section IV, an anomaly detection method using an AE is presented. Performance evaluations are presented in Section V. Finally, the article is concluded in Section VI.

**II. ANOMALY DETECTION**

The detection of anomalies has provided a classic explanation of problems across multiple domains, ranging from scientific observations to financial transactions [23]. We define anomalies as items, events, or observations in the data that deviate significantly from other items, events or observations in terms of behavior so as to arouse suspicion. Anomalies are also referred to as abnormalities, deviants, or outliers in the data mining and the statistics literatures [31]. Several attempts have been made to detect anomalies in order to define an area that reflects the normal behavior; any observations outside the defined area is an anomaly [35].

Supervised deep learning has been widely researched in many fields [32], [36], [37]. However, most supervised methods require a set of labeled datasets of both the normal and the abnormal classes to train a deep supervised binary or a multi-class classifier. Although the performance of the supervised models for detecting anomalies has improved, but in practice they still face some problems. First, it is difficult to define a normal area that contains all possible normal behaviors. Furthermore, the distinction between normal and abnormal behaviors are often unclear, an apparently abnormal observation near the boundary may in fact be normal and vice versa. Secondly, the concept of normal behavior continues to evolve and it might not be possible to adequately represent an existing notion of normal behavior. Thirdly, labeled data is usually difficult and expensive to obtain, which lead to the lack of labeled training samples [35]. Finally, the data usually
includes noise that tends to be close to the real anomalies, making it difficult to identify and delete them.

In contrast to supervised learning, unsupervised learning consists of working with unlabeled data [23], [31], [35]. Unsupervised approaches do not require labeled anomaly data, so they are more suitable for fault diagnosis to learn the underneatth features without the requirement of a labeled dataset [31], [32]. The basic idea behind the unsupervised anomaly detection approach is to find an approximate model capable of capturing the normal behaviors of complex systems. The approximate model could then be used to mark anomalies if the deviation of the predicted behavior of the trained model from the actual observation exceeds a certain threshold. However, it is difficult to converge high dimensional data in the unsupervised technique as it is less accurate than the supervised technique.

Semi-supervised learning deals with a partially labeled dataset to detect anomalies [38], [39]. It has advantages of both supervised and unsupervised learning and good accuracy, even when the dataset is not fully labeled. In this study, we propose a semi-supervised anomaly detection method using AE. The proposed method uses unsupervised learning to learn the best possible representation of data and exploits supervised learning to determine the threshold.

III. ON-SITE NOISE AND PRPD MEASUREMENTS

In this section, we present the on-site noise and PD measurements using an on-line UHF PD monitoring system for GISs [40]. Fig. 1 shows a block diagram that is composed of a GIS, an internal UHF sensor, and a data acquisition system (DAS), for noise and PD measurements. The internal UHF sensor is used with a frequency range of 0.5 GHz to 1.5 GHz and a sensitivity of −14.5 dBm at 5PC. Verification was performed by CIGRE TF 15/33.05 [7].

A. ON-SITE NOISE MEASUREMENTS

Noise was measured for 312 cases using a commercial on-line UHF PD monitoring system for on-site GISs in 6 substations in South Korea and other countries. Fig. 2 shows an example of on-site noise signals for 1000 power cycles and the corresponding 2D representation, where the noises of 1000 power cycles are accumulated to generate the 2D representation, where the noises of 1000 power cycles and 2D representations.

Fig. 2(a) shows the number of noises per 1000 power cycles is illustrated by seven types of PRPD signals for on-site PRPDs with 1000 power cycles and 2D representations.

B. ON-SITE PRPD MEASUREMENTS

For on-site PRPD measurements, 215 cases of on-site PD data were collected from 2003 to 2015 [7]. Figs. 3 and 4 show seven types of PRPD signals for on-site PRPDs with 1000 power cycles and 2D representations.

Figs. 3a and 4a show that crack PDs have a dense distribution in both halves of the cycle, with a maximum amplitude of −50 dBm. Figs. 3b and 4b show that the floating PDs have a sparse distribution in both the halves, with an amplitude of −30 dBm. Figs. 3c and 4c show that the free particle PDs are sparsely distributed in both the halves, but with an amplitude of −50 dBm. Figs. 3d and 4d show void PDs clearly across all bands, with the amplitude varying from −65 to −55 dBm. In Figs. 3e and 4e, POC PDs are observed mostly from 90° to 180°. Figs. 3f and 4f show that POE PDs are densely distributed from 270° to 360°. Figs. 3g and 4g show that the POS PDs mostly occur from 0° to 90° and from 180° to 300°, and with a sparse occurrence from −75 to −50 dBm.

For on-site noise and PRPD measurements, Fig. 5 depicts an example of statistical features using mean(PDs) and max(PDs), where mean(PDs) and max(PDs) are defined as

\[
\text{mean}(\text{PDs}) = \text{mean}(X) = \frac{\sum_{m=1}^{M} x_{m}}{MN} \quad \text{and} \quad \text{max}(\text{PDs}) = \max(X) = \max\{x_{11}, x_{12}, \ldots, x_{MN}\},
\]

respectively. In the case of on-site noise, mean(PDs) and max(PDs) were distributed in [−80, −24] dBm and [−73, −20] dBm, respectively. In the case of on-site noise, mean(PDs) and max(PDs) were distributed in [−80, −24] dBm and [−73, −20] dBm, respectively. For the free particle, mean(PDs) was observed mostly at −73 dBm and the max(PDs) in [−47, −32] dBm. The POC has a range of mean values in [−75, −62] dBm and a range of max values in [−63, −30] dBm. The POE has mean(PDs) and max(PDs) distributed in [−70, −60] dBm and [−60, −35] dBm, respectively. The mean(PDs) and max(PDs) of POS mostly occur in [−80, −73] dBm and [−52, −37] dBm, respectively. In the case of the crack, mean(PDs) occurred in [−80, −60] dBm, and max(PDs) distributed in [−60, −30] dBm. As depicted in Fig. 5, it is...
FIGURE 3. Examples of on-site PRPDs in the GISs for: (a) crack, (b) floating, (c) free particle, (d) void, (e) POC, (f) POE, and (g) POS.

difficult to distinguish between noise and PRPDs, because there is information loss in the process of feature extraction by statistical parameters, such as mean(PDs) and max(PDs).

IV. PROPOSED SCHEME

In this section, we define the anomaly detection problem for PD diagnosis and describe the architecture of the proposed method to detect PRPDs in a GIS. The proposed model employed the autoencoder structure using a training process based on the noise data only and determined the threshold using the noise data and the PRPDs in a semi-supervised manner.

A. PROBLEM FORMULATION

In anomaly detection, we assumed that the training data contained normal data points only, and we identified whether a new sample was an anomaly.

Let $L(X) : \mathbb{R}^{M \times N} \rightarrow \mathbb{R}_{\geq 0}$ denote a distance function mapping input matrix to a positive value that shows how far a sample is from a normal state, where $M$ is the power cycle, $N$ is the phase angle, and $\mathbb{R}_{\geq 0} = \{x \in \mathbb{R} | x \geq 0\}$ is the set of positive real numbers. The higher the value of $L(X)$ the higher is the chance of the corresponding data point being abnormal. For a given threshold value $\varepsilon > 0$, we define the detection accuracy $L(X)$ as the ratio of the correctly detected anomaly samples with $L(X) > \varepsilon$ to the normal ones as $L(X) < \varepsilon$.

Our goal was to learn the score function $L(X)$ and the corresponding threshold $\varepsilon$, to achieve the best detection accuracy of anomalies on the new test data, while minimizing the falsely identified normal sample.

B. AUTOENCODER ARCHITECTURE

An autoencoder is a neural network using a bottleneck structure. The goal of training the autoencoder was to minimize the difference between the reconstructed input and the original input [32]. Given $X \in \mathbb{R}^{M \times N}$, the encoder transforms $X$ to a latent representation $z \in \mathbb{R}^a$ and the decoder is trained to use...
z to reconstruct the original input, where a is the dimension of the latent vector [41].

Fig. 6 shows the detailed structure of the proposed AE. This model consists of two major components: an encoder (for encoding input) and a decoder (for reconstruction). The encoding part transforms the inputs into an internal representation and the internal representation is translated into the output by the decoding part.

For an input, a flattened vector $x_{\text{Flattened}}$ is defined as $x_{\text{Flattened}} = [x_1, x_2, \ldots, x_M]^T$. The output $h^i_e$ of the $i$th encoder layer can be formulated as

$$h^i_e = \psi^i_e(W^i_e h^{i-1}_e + b^i_e)$$

$$= \psi^i_e(W^i_e \phi^{i-1}_e(W^{i-1}_e \phi^{i-2}_e(\ldots(W^2_e \phi^1_e(W^1_e x_{\text{Flattened}} + b^1_e) + b^2_e) + \ldots) + b^{i-1}_e) + b^i_e),$$

where $W^i_e$ and $b^i_e$ are the weight matrix and the bias vector of the $i$th encoding layer ($i = 1, 2, \ldots, L$), respectively, $\phi^i_e$ is the non-linear activation function, and $h^1_e = \psi^1_e(W^1_e x_{\text{Flattened}} + b^1_e)$ is the output of the first encoding layer. By stacking multiple encoding layers, the latent vector $z$ is calculated as

$$z = \psi^L_e(W^L_e h^{L-1}_e + b^L_e).$$

Similarly, the $k$th decoder layer is a fully connected layer and can be calculated as

$$h^k_d = \psi^k_d(W^k_d h^{k-1}_d + b^k_d),$$

where $W^k_d$ and $b^k_d$ are the weight matrix and the bias vector of the $k$th decoding layer ($k = 1, 2, \ldots, K$), respectively, $\psi^i_d$ is the non-linear activation function and $h^1_d = \psi^1_d(W^1_d z + b^1_d)$ is the output of the first decoding layer. Finally, the reconstructed input matrix $\hat{X}$ is calculated from the last decoding layer $h^K_d$, where

$$h^K_d = \psi^K_d(W^K_d h^{K-1}_d + b^K_d).$$

The parameters of the proposed AE model were learnt through the mini-batch $B$ to minimize the loss function, where the parameters included the hyperparameters, weight parameters, and bias parameters. The loss function of the $i$th training data is calculated as

$$\ell(X_i, \hat{X}_i) = \|X_i - \hat{X}_i\|_2^2,$$

where $\hat{X}_i$ denotes the reconstructed sample corresponding to the training sample $X_i$. The total loss $J$ is calculated as

$$J = \frac{1}{|B|} \sum_{i=1}^{|B|} \ell(X_i, \hat{X}_i).$$

Before training, the input dataset in the range of $x_{\text{min}}$ and $x_{\text{max}}$ was normalized to $[0, 1]$, where $x_{\text{min}}$ and $x_{\text{max}}$ were the minimum and maximum values of the original dataset, respectively [32]. The weights and biases were updated based on the gradient information of the total loss. The Adam algorithm [41] was chosen as the gradient descent method because it required that only the first-order gradient be calculated, thus reducing the calculation complexity.

### C. ANOMALY DETECTION USING AUTOENCODER

The proposed AE aimed to find a compact representation of the input data distribution. The AEs are generally data-specific, and their utility is restricted to data that is considerably similar to their training data [43]. For anomaly detection, the AE model was first trained on a noise-only dataset to regenerate noise signals. Thus, it is difficult to reconstruct PRPD signals using the AE model. Here, the reconstruction error was used as an anomaly score to indicate the potential anomaly, because the reconstruction errors of PRPD signals of the proposed AE are larger than those of the noise signals.

After training, we assumed that the trained AE model had learned a good representation of the normal signal pattern, recorded by the UHF sensor. Therefore, the differences between the reconstructed signals and their corresponding inputs could be considered as an important metric for detecting the possible anomalies. An error threshold was identified based on the performance of the model using the test set containing both noise samples and PRPDs to discriminate the PRPDs from the noise data. To determine the threshold, the receiver operating characteristic (ROC) curve with the test set was used in a semi-supervised manner [44]. The threshold $\varepsilon$ was selected by minimizing the distance from the corresponding point on the ROC curve to the ideal threshold, where the true positive rate (TPR) was defined as the probability of detection of the fault data, detected as the fault estimate. The false positive rate (FPR) was defined as the probability of the false alarm that was estimated using the noise data. The ideal threshold had $\text{TPR} = 1$ and $\text{FPR} = 0$. Fig. 7 shows a semi-supervised learning method for anomaly detection.
V. PERFORMANCE EVALUATION
This section presents the performance evaluation using on-site noise and PRPD measurements. Table 1 shows the number of samples for on-site noise and PRPDs in GISs, where seven types of PRPDs such as crack, floating, free particle, POC, POE, POS, and void PDs, are considered. Each experiment was performed with $M = 1000$ power cycles and $N = 256$ phase angles. Noise was measured using the UHF sensor in on-site fields. For semi-supervised learning, we divided the experimental dataset into training, validation, test, and new test sets [44]. The training set was 80% of the noise data and the validation set was 10% of the noise data. To determine the threshold, we used 5% of the noise data and 50% of the PRPD data for the test set. Then, the detection accuracy was calculated using the new test set, which consisted of 5% of the noise data and 50% of the PRPD data.

For AE, we deployed the encoder and decoder as plain and fully connected neural networks. Each layer was followed by a rectified linear unit (ReLU) activation with no batch normalization. To acquire the optimized hyperparameters, such as the batch size, number of layers, latent vector size, etc., the proposed model was trained using several combinations of parameters, and the best combination was selected. The AE model with $L = 7$ layers, a latent vector size $= 32$, a batch size with $|B| = 16$ and a learning rate of 0.0001 was chosen. The details of the proposed AE are listed in Table 2, where the total parameters for the training set are 131,408,736.

Reconstruction loss values were used to detect faults by selecting an appropriate detection threshold. This is because the AE model was trained using noise data only. It was expected that the loss values of the noise data would be lower than those of the fault data due to the differences in the data distribution.

Fig. 8 depicts the ROC curve of the proposed AE model with different thresholds using the test set. The thresholds are selected between 0 and 0.45, considering the range of reconstruction errors of the samples in the test set, where TPRs and FPRs are calculated using thresholds. The threshold corresponding to the point, which is the closest to the ideal point of the ROC curve, is determined by $\epsilon = 0.023$ for the new test set. The area under the curve (AUC) was calculated as 0.76 with the determined threshold [45].
FIGURE 8. ROC curve of the proposed AE model on the test set.

Fig. 9 shows the reconstruction error on the new test set, containing 50% fault dataset and 5% noise data. From these histograms, most noise samples are not detected falsely as fault data, whereas in some fault data, the reconstruction loss is higher than the threshold. To analyze the errors of the proposed AE, Fig. 10 shows the scatter diagram of loss values on the new test set with threshold $\varepsilon = 0.023$ and data visualization on errors for noise, crack, and POS. For anomaly detection, the proposed AE detects most of the PRPDs from noise samples, as shown in Fig. 10a. This is because the features are learned by a small number of nodes in the middle layer of the proposed AE using the on-site noise data for GISs to effectively reconstruct the input. Amplitudes of the noise sample in Fig. 10b are higher than those of the noise data in Fig. 2a. As shown in Figs. 10c and 9d, PRPDs for crack and POS have lower amplitudes compared to Figs. 3a and 3g, respectively. This is because some kinds of crack and POS were detected as noise samples.

Table 3 lists the accuracy of the AE model compared to an OCSVM [46]. In our experiment, the OCSVM classifier used the RBF kernel, $\mu = 0.3$, and $\gamma = \text{‘scale’}$ after multiple trials using $\mu \in (0, 1]$ to find the best model [34]. The proposed AE model achieved 86.75% accuracy on the new test set and was approximately 40.25% higher than OCSVM. This is because the proposed AE could extract meaningful features from the noise data and reliably detect the PRPD. The proposed AE outperformed in the particle and the POS types, and successfully detected 100% and 50% of the samples, respectively, whereas OCSVM could not detect the particle and the POS types in our experiment. Only in the POE type, OCSVM achieved 100% accuracy and was higher than our method by 20% in terms of accuracy. With the other classes, the proposed AE model was more accurate than OCSVM, with approximately 3.33%, 18.75%, 42.42%, and 73.81% differences for floating, crack, POC, and void, respectively. Notably, the performance of the proposed AE was achieved with much fewer false positive detection cases compared to OCSVM, where only 6.25% of the noise samples were falsely recognized as fault data.

To understand better what the model learned, we analyzed the features of the input and the latent vectors. Fig. 11 shows the t-distributed stochastic neighbor embedding (t-SNE) representation to visualize a set of inputs composed of both noise and PRPDs, and their output extracted from the latent layer [47]. Here, t-SNE helped in reducing the dimensions of the data, from a multi-dimensional vector to only the top 2 components with maximum variation and visualized them such that similar objects were transformed into nearby points. Fig. 11a shows that numerous fault data are very close to the noise data and hence difficult to classify accurately. As shown
TABLE 3. The accuracy comparison using the new test set.

| Class     | Overall | Noise | Crack | Floating | Particle | POC | POE | POS | Void |
|-----------|---------|-------|-------|----------|----------|-----|-----|-----|------|
| OCSVM     | 46.25   | 38.10 | 56.25 | 96.67    | 0        | 57.58 | 100 | 0   | 21.43 |
| AE        | 86.75   | 93.75 | 75    | 100      | 100      | 100  | 80  | 50  | 95.24 |

FIGURE 10. (a) Scatter diagram of the loss values on the new test set after PD detection with threshold $\epsilon = 0.023$ and the visualization of the typical samples detected incorrectly from: (b) noise, (c) crack, and (d) POS.

FIGURE 11. t-distributed stochastic neighbor embedding (t-SNE) representation of 123 samples of the new test set at: (a) the input data and (b) the latent vector.

in Fig. 11b the floating, free particle, POC, and void types are separated into noise after the encoding process because the proposed model extracted the meaningful features for reconstruction. It is seen that some samples of the crack and the POS types are close to the noise samples, which affected the low detection performance of the proposed AE, as listed in Table 3. Also, the POE slightly overlaps with the noise sample and hence the proposed AE detection performance for POE was 80%.

VI. CONCLUSION
In this article, we proposed an AE-based anomaly detection method to detect PRPDs in GISs. The proposed AE was solely trained with noise data, and the threshold was determined using noise and PRPD data in a semi-supervised manner. The proposed AE has an advantage of feature extraction, based on the on-site noise data for GISs. It was verified based on the on-site noise and PD measurements, using an online UHF PD monitoring system for GISs. The on-site PD data included seven types of faults, such as the crack, floating, free particle, POC, POE, POS, and void. The experimental results revealed that the proposed AE achieved a detection accuracy of 86.75% and had a 40.25% higher detection per-
formance than the OC-SVM. The proposed AE can be applied to offline anomaly detection based on the noise and PRPD measurements, using an offline system. In future studies, we intend to design artificial cells for analyzing the PRPD patterns considering various severities of faults and conduct further verifications of the proposed method for all severity levels.

REFERENCES

[1] P. Shah and K. Gehring, “Smart solutions to power the 21st century: Managing assets today for a better grid tomorrow,” IEEE Power Energy Mag., vol. 14, no. 2, pp. 64–68, Mar. 2016.

[2] D. Ma, X. Hu, H. Zhang, Q. Sun, and X. Xie, “A hierarchical event detection method based on spectral theory of multidimensional matrix for power system,” IEEE Trans. Syst., Man, Cybern. Syst., early access, Aug. 9, 2019, doi: 10.1109/TSMC.2019.2931316.

[3] W. Rui, S. Qiuye, M. Dazhong, and H. Xuguang, “Line impedance cooperation region identification method for grid-tied inverters under weak grids,” IEEE Trans. Smart Grid, vol. 11, no. 4, pp. 2856–2866, Jul. 2020.

[4] R. Wang, Q. Sun, P. Zhang, Y. Gui, D. Qin, and P. Wang, “Reduced-order transfer function model of the droop-controlled inverter via Jordan continued-fraction expansion,” IEEE Trans. Energy Convers., early access, Mar. 12, 2020, doi: 10.1109/TEC.2020.2980033.

[5] H. Ma, T. K. Saha, C. Ekanayake, and D. Martin, “Smart transformer for smart grid—Intelligent framework and techniques for power transformer asset management,” IEEE Trans. Smart Grid, vol. 6, no. 2, pp. 1026–1034, Mar. 2015.

[6] S. Li, Y. Nie, and J. Li, “Condition monitoring and diagnosis of power equipment: Review and prospective,” High Voltage, vol. 2, no. 2, pp. 82–91, May 2017.

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

[8] U. Schichtler, W. Koltunowicz, D. Gautschi, A. Girodet, H. Hama, K. Juhere, J. Lopez-Roldan, S. Okabe, S. Neuhold, C. Neumann, J. Pearson, R. Pietsch, U. Riechert, and S. Tenbohlen, “UHF partial discharge detection system for GIS: Application guide for sensitivity verification,” in Proc. IET Sci., Meas. Technol., vol. 14, no. 2, pp. 64–68, Mar. 2016.

[9] Q. Khan, S. S. Refaat, A. Abu-Rub, and H. A. Toliyat, “Partial discharge detection and diagnosis in gas insulated switchgear: State of the art,” IEEE Elect. Insul. Mag., vol. 35, no. 4, pp. 16–33, Jul. 2019.

[10] M. Wu, H. Cao, J. Cao, H.-L. Nguyen, J. B. Gomes, and R. A. Sánchez, Ed. Rijeka, Croatia: InTech, 2015, pp. 102–109.

[11] X.-J. Shao, and J.-N. Zhang, “Classification and separation of partial discharge ultra-high-frequency signals in a 252 kV gas insulated substation using cumulative energy technique,” IET Sci., Meas. Technol., vol. 10, no. 4, pp. 316–326, Jul. 2016.
[37] E. L. Paula, M. Ladeira, R. N. Carvalho, and T. Marzagao, “Deep learning anomaly detection as support fraud investigation in Brazilian exports and anti-money laundering,” in Proc. 15th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA), Anaheim, CA, USA, Dec. 2016, pp. 954–960.

[38] I. Monroy, R. Benitez, G. Escudero, and M. Graells, “A semi-supervised approach to fault diagnosis for chemical processes,” Comput. Chem. Eng., vol. 34, no. 5, pp. 631–642, May 2010.

[39] J. Ma and J. Jiang, “Semisupervised classification for fault diagnosis in nuclear power plants,” Nucl. Eng. Technol., vol. 47, no. 2, pp. 176–186, Mar. 2015.

[40] Hysoung Corporation. Hysoung Intelligent Condition Monitoring System Catalogue. Accessed: Jul. 18, 2020. [Online]. Available: http://hysoung.takout.co.kr/files/catalogue/HICMS/ENG.pdf

[41] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, and A. van den Hengel, “Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection,” Aug. 2019, arXiv:1904.02639. [Online]. Available: http://arxiv.org/abs/1904.02639

[42] D. P. Kingma and J. Ba. “Adam: A method for stochastic optimization,” Jan. 2017, arXiv:1412.6980. [Online]. Available: http://arxiv.org/abs/1412.6980

[43] N. Purkait, Hands-On Neural Networks With Keras: Design and Create Neural Networks Using Deep Learning and Artificial Intelligence Principles, 1st ed. Birmingham, U.K.: Packt Publishing Ltd., 2019. [Online]. Available: http://shorturl.at/tMTZ1

[44] Y. Pan, F. Sun, Z. Teng, J. White, D. C. Schmidt, J. Staples, and L. Krause, “Detecting Web attacks with end-to-end deep learning,” J. Internet Services Appl., vol. 10, no. 1, pp. 1–22, Dec. 2019.

[45] D. Abati, A. Porrello, S. Calderara, and R. Cucchiara, “Latent space autoregression for novelty detection,” Mar. 2019, arXiv:1807.01653. [Online]. Available: http://arxiv.org/abs/1807.01653

[46] B. Schölkopf, R. C. Williamson, A. J. Smola, J. Shawe-Taylor, and J. C. Platt, “Support vector method for novelty detection,” in Advances in Neural Information Processing Systems, S. A. Solla, T. K. Leen, and K. Müller, Eds. Cambridge, MA, USA: MIT Press, 2000, pp. 582–588.

[47] 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.

---

NGOC-DIEM TRAN THI was born in Quang Ngai, Vietnam, in 1996. She received the bachelor’s degree in statistics from Ton Duc Thang University, Ho Chi Minh, Vietnam, in 2018. She is currently pursuing the master’s degree with the Information Technology Convergence Laboratory, Department of Electronic Engineering, Myongji University, South Korea, under the supervision by Prof. Y.-H. Kim. Her research interests include statistics, data analysis, deep learning, and application of deep learning to discover potential of data science.

THE-DUONG DO received the B.Eng. degree in mechatronics from the Hanoi University of Science and Technology (HUST), 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 comprise deep learning applications for fault diagnosis, generative adversarial nets, and automotive radar signal processing.

JAE-RYONG JUNG was born in Busan, South Korea, in 1976. He received the B.Sc. and M.Sc. degrees from Busan National University, Busan, in 1999 and 2001, respectively. Since then, he has been working with the Research and Development Center, Hysoung Corporation, where he is the Chief Researcher. His research interests are partial discharge diagnostics system for power apparatus and asset management solution for substations. He is also a Korea Representative Member of CIGRE SC B3.

HYANGEUN JO was born in Tongyoung, South Korea, in 1987. She received the B.Sc., M.Sc., and Ph.D. degrees from Korea Maritime and Ocean University, Busan, South Korea, in 2010, 2012, and 2016, respectively. Since then, she has been working with the Research and Development Center, Hysoung Corporation. Her research interests are partial discharge of GIS and dissolved gas analysis diagnosis method of power transformers.

YONG-HWA KIM (Member, IEEE) received the B.S. degree in electrical engineering and the Ph.D. degree in electrical and computer engineering from Seoul National University, Seoul, South Korea, in 2001 and 2007, respectively. From 2007 to 2011, he was a Senior Researcher with the Korea Electrotechnology Research Institute (KERI), Gyeonggi-do, 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. His general research interests include communication systems, fault diagnosis, and digital signal processing. He is currently particularly interested in artificial intelligence for communications, radar systems, and smart grid.