

# Identification of Partial Discharge Defects Based on Deep Learning Method

Lian Duan<sup>1</sup>, Jun Hu<sup>1</sup>, Gen Zhao<sup>1</sup>, Kunjin Chen<sup>1</sup>, Jinliang He<sup>1</sup>, *Fellow, IEEE*, and Shan X. Wang, *Fellow, IEEE*

**Abstract**—Since repetitive partial discharge (PD) leads to insulation breakdown, it is one of the most critical defects that affect operation life of electrical equipment. In this paper, four kinds of PD defects are identified with deep learning (DL) method according to the current waveforms. A modified IEC-60270 experiment platform with ultra-high frequency (UHF) and current probe is built to acquire PD current waveforms and their corresponding detecting pulse current and UHF pulse signal. Fourier transform, principle component analysis, and t-distributed stochastic neighbor embedding methods are applied to visualize the data set, which proves the feasibility of classifying the PD current waveform. Two basic parts of this DL framework are sparse autoencoder layer and softmax layer, the former extracting features of the input signal and the latter operating as the classifier. Hyper-parameters of the network such as sparsity, activation function, number of hidden nodes, and network depth were discussed. The final classifying accuracy of the proposed method is up to 99.7%, that is much better than the traditional identifying method. A comprehensive blind test is designed to prove the general validity and robustness of the proposed model.

**Index Terms**—Partial discharge, deep learning, sparse autoencoder, softmax, feature extraction.

## I. INTRODUCTION

LONG with the development of the smart grid, several approaches have been proposed to make the electrical equipment smarter and smarter. One of the prerequisite of the smart equipment is reliable health monitoring system. Therefore, once a certain fault occurs, the equipment is able to identify the fault type by itself and react to it afterwards. Partial discharge (PD) is an early indicator on insulation breakdown that will cause catastrophic failure on the power system, so PD identification is a significant approach to monitor the operating status of the electrical equipment [1]. PD has been studied by numerous researchers, the area of which includes PD detection, identification and location.

Manuscript received October 10, 2018; revised January 23, 2019 and March 19, 2019; accepted April 3, 2019. Date of publication April 11, 2019; date of current version July 23, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 51720105004 and in part by the Research Project of State Grid Corporation of China under Grant 5202011600UJ. Paper no. TPWRD-01199-2018. (*Corresponding author: Jun Hu.*)

L. Duan, J. Hu, G. Zhao, K. Chen, and J. He are with the Department of Electrical Engineering, Tsinghua University, Beijing 100084, China (e-mail: duanl15@mails.tsinghua.edu.cn; hjun@tsinghua.edu.cn; zhaogthu@gmail.com; yalickj@163.com; hejl@tsinghua.edu.cn).

S. X. Wang is with the Center for Magnetic Nanotechnology, Stanford University, Stanford, CA 94305 USA (e-mail: sxwang@stanford.edu).

Digital Object Identifier 10.1109/TPWRD.2019.2910583

Several PD detection methods have been proposed such as pulse current method (also referred as electrical research association (ERA) method) [2–3], ultra-high frequency (UHF) method [4], [5], dissolved gas analysis (DGA) method [6], acoustic method [7] and optical method [8]. Among all of them, ERA and UHF are the most widely used method. ERA is the earliest PD detection approach suggest by IEC-60270 standard. It is an intrusive measurement so it can only be used as offline PD test with embedded coupling capacitance and detecting impedance. By using this method, the actual charge quantity of discharge could be measured, although it must be calibrated by complicated structure. In order to improve the anti-interference ability of ERA, UHF method was proposed by Hampton in 1988. Thanks to the non-intrusive and high bandwidth feature, it has become the most popular method to detect and locate PD. Since UHF signal is only related to the change rate of the discharge current, it cannot provide actual discharge quantity [9]. Thus this method cannot indicate the deterioration condition of the insulation. Although several researchers try to find the relationship between the UHF peak value and discharge quantity [10], [11], there is no universal calibration standard to precisely estimate the actual discharge level.

According to operating experience, different PD defects will cause various damage to electrical apparatus. Typical procedure of PD identification usually contains two steps, feature extraction and classification. In the first step, feature is generally extracted from time domain waveforms obtained using ERA and UHF method. Previous studies have indicated that discharge amplitude (Q), repetitive times (N) and phase ( $\varphi$ ) are closely related to PD type [12], which is called phase resolved PD analysis (PRPD). Traditionally, a 3D or split 2D  $\varphi$ -N-Q charge was design as the feature [13]. U. Schichler also uses NoDi\* pattern as the input feature of the classification algorithm in DC voltage scenario [15], [16]. As for the second step, recently, machine learning techniques (such as neural network, expert system and fuzzy system, e.g.,) have increasingly provided efficient methods for PD detection and recognition [17]–[19]. Moreover, since sparse auto-encoder (SAE) was proposed by Hinton in 2006 [20], machine learning has stepped to the so-called ‘deep world’. In [12], this deep learning method is applied to assess PD severity of gas insulated switchgear (GIS). A stacked sparse auto-encoder (SSAE) is proposed to realize feature extraction from the middle layer with a small number of nodes. The features extracted from PD data are then fed into a soft-max classifier to be classified into one of four defined PD severity states.

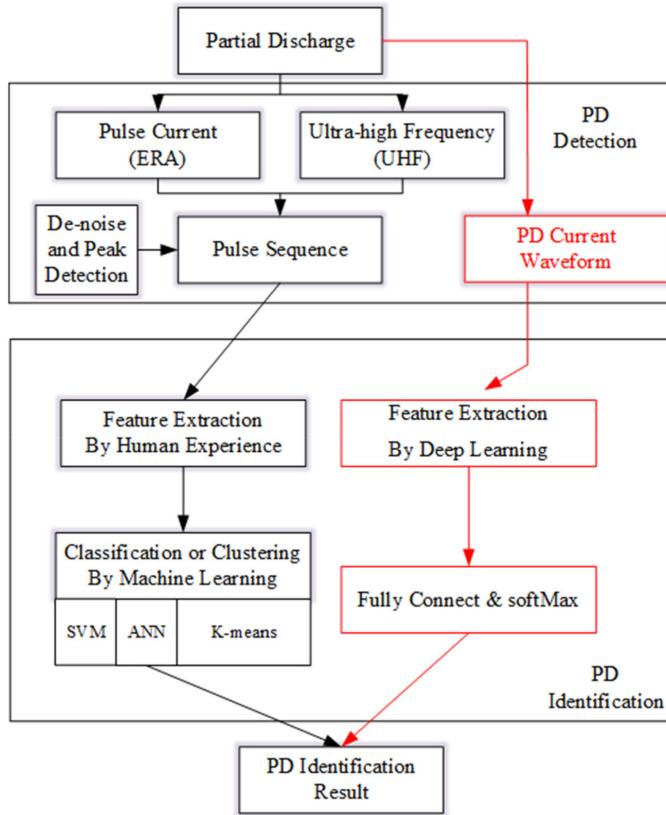


Fig. 1. Previous flowchart of PD identification (black) and its modification in this paper (red).

According to previous summary, the process of determining PD type is listed in Fig. 1. However, there are several disadvantages in this procedure. Foremost, the pulse sequence is detected by ERA and UHF, yet none of them contains the whole information of the inner of discharge progress, despite the shortcoming listed above. Due to the lack of information, PRPD must be done to construct distinguishable input features. Furthermore, deep learning is proposed to do the job of feature extraction automatically only by machine itself rather than human experience. Thus, it will be a redundancy if PRPD and deep learning were applied simultaneously.

An intelligent framework for power transformer asset management is proposed in [21], including multiple online sensor measurements, data preprocessing unit (made of de-noising and feature extraction), pattern recognition, multi-dimensional data fusion and transformer condition assessment by artificial intelligence (AI) algorithm. In this paper, PD identification framework is designed similarly. In order to solve the contradiction between PRPD and deep learning method, PD current is suggested to be used as input signal verified by the traditional detecting method both ERA and UHF. All the waveforms are acquired by the self-made test platform, which is regarded as the sensor measurement. Deep learning method is applied to distinguish four kinds of artificial PD defects, using SAE to extract the feature automatically by the algorithm and softmax to identify PD type, corresponding to data preprocessing unit and pattern recognition respectively. In the following sections, preparation of PD

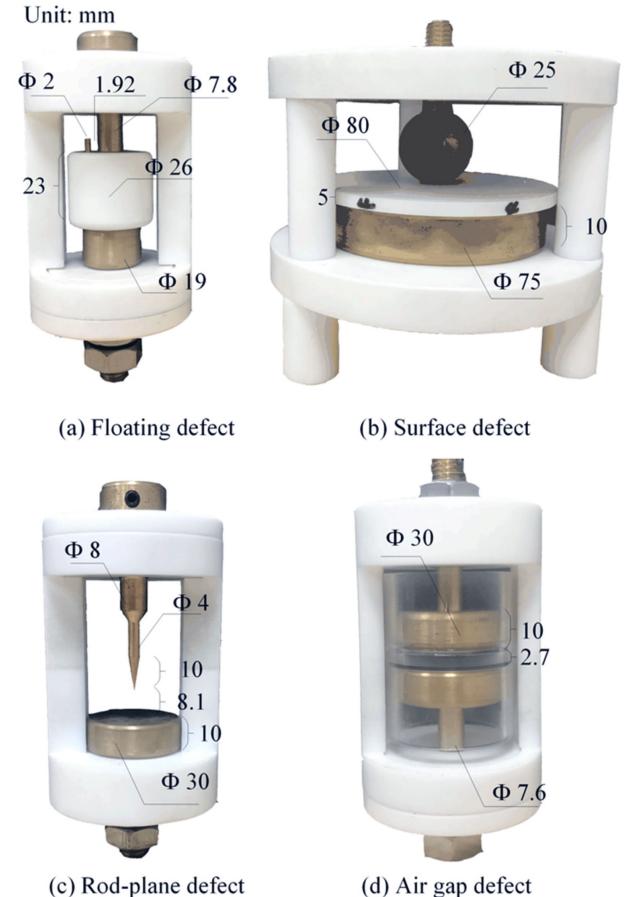


Fig. 2. Four kinds of artificial PD defects.

defect, data acquiring system, deep learning algorithm and its hyper-parameters and a blind test will be discussed step by step.

## II. DATA ACQUISITION AND FEASIBILITY OF CLASSIFICATION

### A. PD Defects and Data Acquiring System Design

*1) Preparation of PD Defects Model:* Typical defects in electrical equipment such as cubic switchgear, dry type reactor includes metallic protrusion, free metal particle, contamination on the surface of the spacer and floating potential [22]. Four kinds of defects are made to simulated the actual PD discharges, which are floating, surface, rod-plane and air gap discharge, specific parameters of which are illustrated in Fig. 2, respectively. All the brim of the copper electrode, except the tip of the rod, is chamfered.

Floating discharge happens when a free metal particle floats in the high electrical field. Therefore, it will gain a certain potential because of the capacitance. This process will cause a distortion in electric field thus bringing PD discharge. The structure of this defect is demonstrated in Fig. 2(a). The surrounding bracket and the insulator in the middle are both made of polytetrafluoroethylene (PTFE). When apply a high voltage by the upper and lower electrode, this small bar, 1.92 mm away from the upper electrode, will get a floating potential.

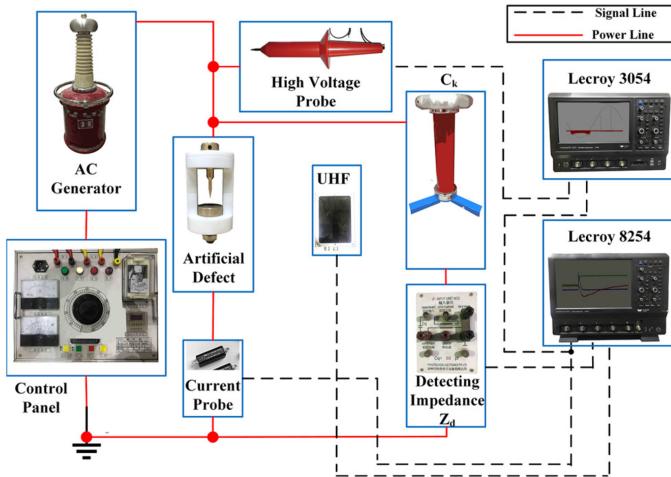


Fig. 3. Test layout.

Surface discharge generally happens at surface of the insulator such as the insulating cardboard place in the oil. CIGRE Method II suggests a typical sphere-plate electrode system, shown in Fig. 2(b). A 5 mm high PTFE is used as the surface. The surrounding framework is also made of PTFE.

Rod-plate electrode model simulates the condition when metal protrusion or cutting burr was place in an electrical field. Shown as Fig. 2(c), the upper rod is made of copper and the distance between the tip of it and the lower electrode is 8.1 mm.

Air gap discharge will take place when there is air bubble or cell defect in the oil or solid insulating material. In this artificial model, a 2.7 mm air gap is sealed within the plexiglass, clamped by two copper electrode.

2) *Experimental Platform:* The platform is developed from the method introduced in IEC-60270 and the circuit is shown in Fig. 3, from which 3 kinds of signal are acquired. PD current waveform is measured by CT-1 probe manufactured by Tek company located at the grounded end. Its frequency response range is 25 kHz to 1 GHz and the minimum response time is 0.35 ns. As introduced in IEC-60270 standard, another detection branch is paralleled by the artificial PD defect.  $C_k$  is the coupling capacitor with a capacity of 1000 pF and  $Z_d$  is the detection impedance. UHF sensor made by Tsinghua University is placed 10 cm away from the PD defect. 2 Lecroy oscilloscopes are used to acquire these three kinds of PD signals and applied AC voltage. Lecroy Wavesurfer 3054, which has 500 MHz bandwidth and 4 GS/s sampling rate, is used to record 1 cycle (20 ms) voltage waveform and corresponding PD current waveform. Single PD current waveform, pulse current signal (ERA) and UHF detecting signal are collected by Lecroy Wavesurfer 8254 with a bandwidth of 2.5 GHz and sampling rate of 20 GS/s.

#### B. Glance of PD Current Waveform and Traditional PD Detection Signal

Lecroy 3054 is used to record voltage and PD current waveform in one cycle, 20 ms, as shown in Fig. 4 (a), (c), (e), and (g). Sampling frequency is set as 50 MS/s, which is a proper value

that is high enough to get each PD pulse with the storage depth that is able to be provided by the oscilloscope.

PD signals acquired by CT1, ERA method and UHF probe are stored in Lecroy 8254 with 1 GS/s sampling rate in a time interval from  $-10 \mu s$  to  $10 \mu s$ . The rising time of typical PD current is several tens of nanosecond. Thus this configuration will be sufficient for this scenario. Fig. 4 (b), (d), (f) and (h) display 4 kinds of PD current waveforms and corresponding detecting signals respectively. By adjusting the voltage level applied on the PD defect, thousands of PD current waveforms are collected as the data set. Voltage amplitude and corresponding number of data are listed in Table I.

### III. PRE-ANALYSIS AND VISUALIZATION OF THE ACQUIRED WAVEFORM DATA SET

#### A. FFT Plot Aggregation

Before the different kinds of data are learnt by the computer algorithm, there must be a pre-analysis to make sure these four kinds of data are distinguishable. It is very tough to classify four kinds of PD current in Fig. 4(b), (d), (f), (h) at one glance because they look like very similar to each other. All of them are generally one pulse with a sharp rising time and an oscillation tail. To make it clear, waveforms in the time domain are transformed to frequency domain. Because of the randomness of PD phenomenon, all the waveforms in one kind are aggregated together in one plot to magnify the characteristic, as shown in Fig. 5.

According to the data glance in second B, it is enough to intercept the waveform data in a  $4 \mu s$  interval, from  $-2 \mu s$  to  $2 \mu s$ , for later use. In this way, considering the sampling rate of 1 GS/s, the highest frequency will be 500 MHz and the frequency step is 125 kHz. Whole magnitude of the waveform is normalized to the range of 0 to 1. It can be seen from Fig. 5 that FFT magnitude of all the four kinds of PD currents are high at the frequency below 10 MHz, but the distribution of each is rather different. At the frequency range from 10 MHz to 100 MHz, surface and rod-plane discharge have only one prominent frequency at different value while floating and air gap discharge have 2 and 3 notable peak values respectively. When the frequency goes higher than 100 MHz, the major information is brought by noise. The frequency of the notable peak in Fig. 5 is listed in Table II.

FFT analysis in Fig. 5 gives an intuitive indication that four kinds of waveform are different from each other. However, the main purpose of this paper is to design a deep learning network that can automatically extract feature from the original data. FFT data is still a kind of man-made feature, so extra work should be done to show the distance among each PD current waveform.

#### B. Visualization Based on Principle Component Analysis (PCA)

PCA is often used to reduce dimension of input data, the objective of which is to linearly map a high dimension vector to a low one. In order to better visualize the data set, first 3 PCA components are selected to stand for the 3 dimensions for a 3-D plot in Fig. 6. As indicated, only rod-plane data could be

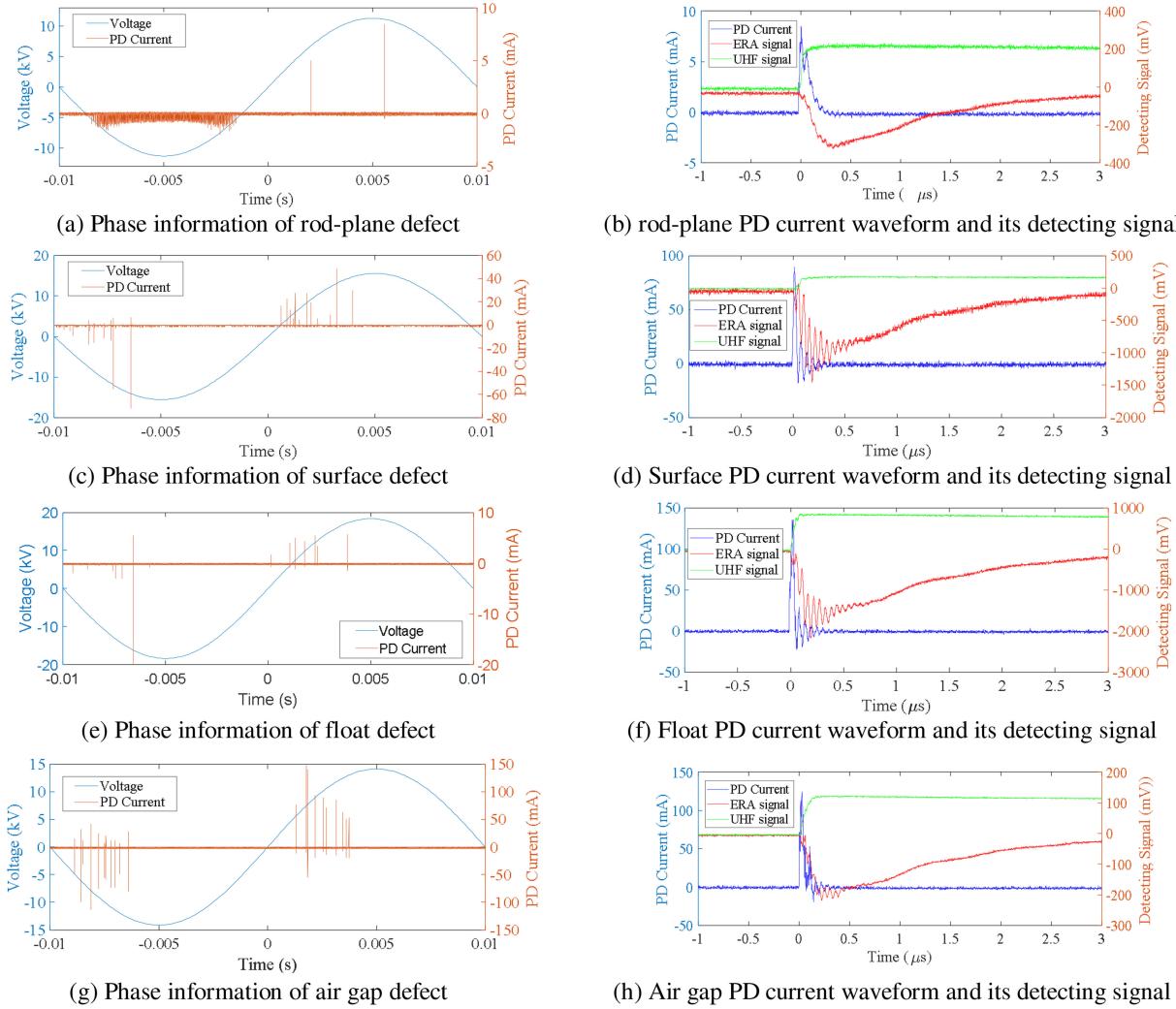


Fig. 4. Different PD signals at a glance.

TABLE I  
AMOUNT OF WAVEFORMS IN THE DATA SET

PD Type	Voltage amplitude (kV)	Number of waveforms
Air Gap	9, 10, 11	1884 (750, 528, 606 respectively)
Float	11.6, 13, 14	1581 (512, 527, 542 respectively)
Rod-plane	4.5, 5.6, 6	1597 (516, 526, 555 respectively)
Surface	9, 11, 12.5	2206 (427, 631, 1103 respectively)

separated by the four kinds of data clearly. At the edge of air gap data, float data tangles with surface data, which makes it difficult to prove the feasibility of classifying the four discharge groups.

### C. Visualization Based on T-Distributed Stochastic Neighbor Embedding (T-SNE)

While PCA is a linear algorithm that cannot explain the polynomial structure among features, T-SNE is able to find out the local feature based on the probability distribution on the

neighborhood graph. The only shortcoming of T-SNE is time complexity. When processing millions of data sample, it could cost T-SNE couple of hours while PCA could finish it in minutes. In this scenario, Barnes-Hut algorithm has the ability to visualize all the 7268 samples in 5 minutes, which is acceptable. When applying this algorithm, hyper-parameters, perplexity and exaggeration, of T-SNE are set to 75 and 100 respectively. Fig. 7 displays the result of T-SNE in different angles. It can be clearly observed that even the air gap, rod-plane and surface data that tangle with each other in Fig. 6 can be separated perfectly. Also, each big group can be classified by several small groups, which means discharge current of a single defect have diversity. One thing to notice is that there are a very small group with all the four kinds of data on the right side. It should be the noise data alone with no PD pulses in them because of mis-trigger of the oscilloscope.

## IV. FRAMEWORK OF CLASSIFICATION

### A. Sparse Autoencoder for Feature Extraction

Sparse autoencoder is a useful tool to extract feature based on deep learning idea [23], which has three components, visible

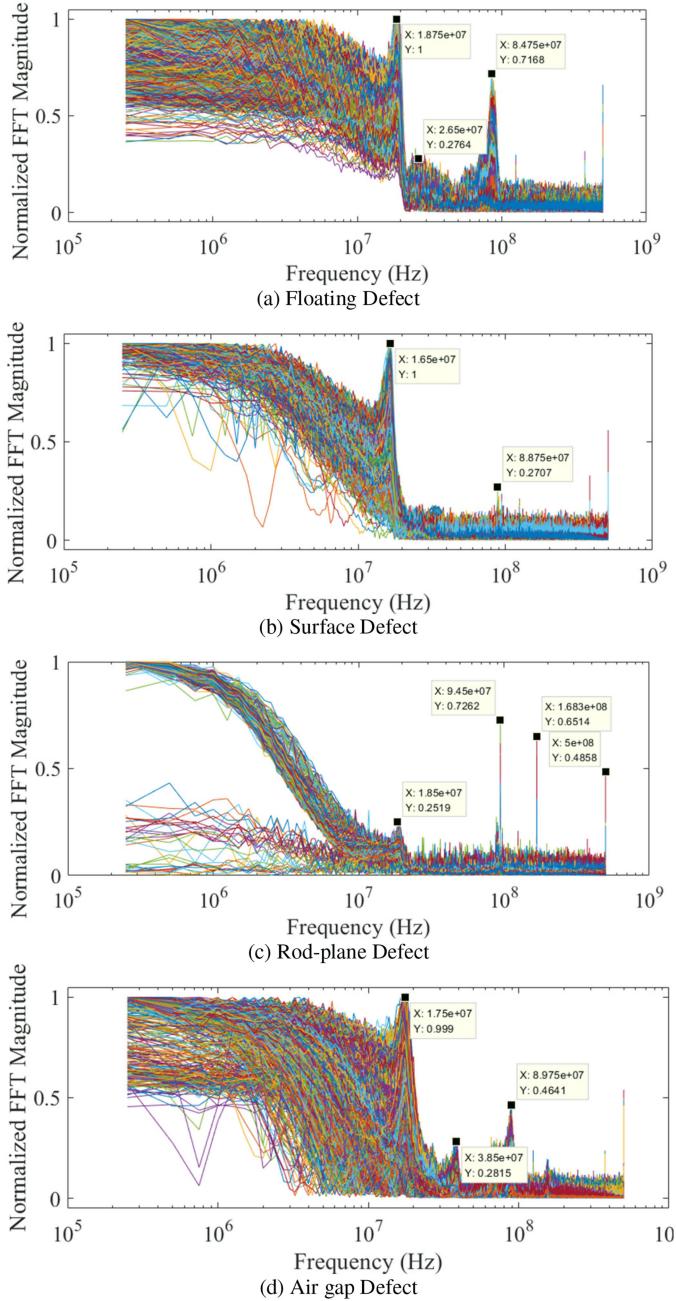


Fig. 5. Aggregate of FFT plot of all the acquired data.

TABLE II  
NOTABLE PEAKS IN THE FFT PLOT

PD Type	Frequencies of Notable peaks (MHz)
Air Gap	17.5, 38.5, 89.75
Float	18.75, 26.5, 84.75
Rod-plane	18.5, 94.5, 168.3, 500
Surface	16.5, 88.75

layer, hidden layer and reconstruction layer, respectively. Generally, the size of hidden layer is smaller than that of the input data. The unsupervised training process tries to make sure that the output vector is the same as the input one. In this way, the

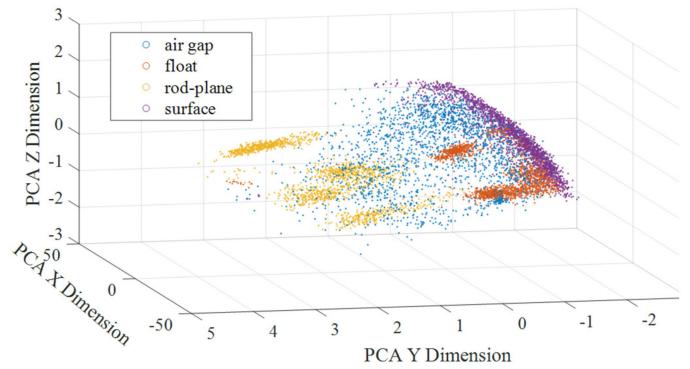


Fig. 6. PCA visualization of the four-group data.

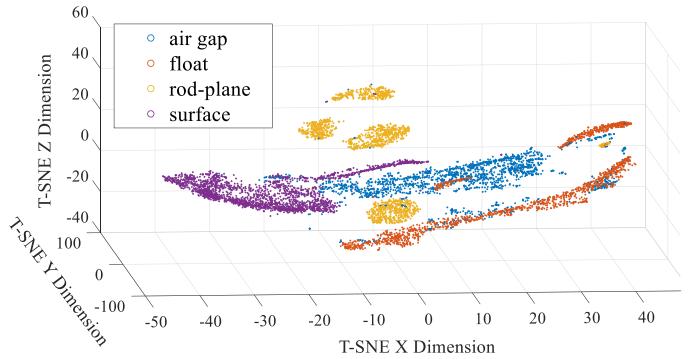


Fig. 7. T-sne visualization of the four-group data.

hidden layer will successfully extract the compressed feature representing the original input signal.

Therefore, only half of the autoencoder, input layer and hidden layer, will be used during application and the feature, output of the autoencoder, will be sent to the following layer which generally is a classification layer. In [23], a SAE is used to extract features of the various types of Ferroresonance overvoltage.

Given a set of input data  $\{x^{(1)}, x^{(2)}, \dots, x^{(i)}\}, x^{(k)} \in \mathbf{R}^n$ , they will be mapped to the hidden layer  $h_l$ ,  $l = 1, 2, \dots, m$  through equation (1)

$$h(x) = f(\mathbf{W}_1 x + \mathbf{b}_1) \quad (1)$$

Where  $\mathbf{W}_1 \in \mathbf{R}^{m \times n}$  is the weight matrix connecting the visible and hidden layer and  $\mathbf{b}_1 \in \mathbf{R}^m$  is the bias vector. In equation (1), function  $f()$  is called the activation function. Typically, it is sigmoid function which is defined as

$$f(a) = 1 / (1 + \exp(-a)) \quad (2)$$

There are also other activation functions, such as tanh, ReLu, ELU and PReLU. Evaluation of the effect of these activation function will be discussed in Section IV. After encoding, the second step, decoding, is completed by the reconstructing layer with (3).

$$\hat{x} = f(\mathbf{W}_2 h(x) + \mathbf{b}_2) \quad (3)$$

Where  $\mathbf{W}_2 \in \mathbf{R}^{n \times m}$  and  $\mathbf{b}_2 \in \mathbf{R}^n$  are also called the weight matrix and bias vector associating output layer with hidden layer.

The accuracy of a simple autoencoder (AE) is assessed by the cost function defined as follow.

$$J_{AE} = \frac{1}{2k} \sum_{i=1}^k \left\| \hat{x}^{(i)} - x^{(i)} \right\|^2 + \frac{\lambda}{2} \sum_{j=1}^2 \|\mathbf{W}_i\|^2 \quad (4)$$

Where the first term calculates the sum-of-squares error of the input and output vector and the second term is the regularization term preventing the AE from overfitting.  $\lambda$  is designed to control the relative magnitude of the two terms. Based on simple AE, SAE introduce sparsity penalty term and the cost function is modified as (5)

$$J_{SAE} = J_{AE} + \beta \sum_{i=1}^m KL(\rho || \tilde{\rho}_i) \quad (5)$$

Where  $\beta$  is the coefficient which controls the weight of sparsity penalty [24] and  $KL(\rho || \tilde{\rho}_i)$  is the Kullback-Leibler divergence between  $\tilde{\rho}_i$ , the average activation of hidden node  $j$  with respect to all input data, and  $\rho$ , the sparsity parameter with a considerably small value, which is defined as follow [24].

$$KL(\rho || \tilde{\rho}_i) = \rho \log \left( \frac{\rho}{\tilde{\rho}_i} \right) + (1 - \rho) \log \left( \frac{1 - \rho}{1 - \tilde{\rho}_i} \right) \quad (6)$$

In a simple way, training a SAE means the optimization of  $J_{SAE}$ . When  $J_{SAE}$  is optimized to an expected small value or a certain number of iterations is reached, the training process will be completed. By using (1), feature of the input vector  $x$  will be extracted automatically.

### B. Softmax for Classification

Softmax classifier is based on softmax regression model, which is a generalization of logistic regression model and could deal with multi-class problems [25]. When softmax is applied to a  $q$ -class classification problem, the output vector of the classifier is made up with all  $q$  probabilities. Thus, the input vector will be divided into the category with the high probabilities. Given a group of labeled input vector  $\{x^{(1)}, x^{(2)}, \dots, x^{(k)}\}$ ,  $x^{(i)} \in \mathbf{R}^n$  and  $y^{(i)}$  being the corresponding label, the probability that  $x^{(1)}$  belongs to a certain class  $j$ ,  $j = 1, 2, \dots, q$  can be calculated as follow.

$$P(y^{(i)} = j | x^{(i)}) = \frac{e^{\theta_j x^{(i)}}}{\sum_{j=1}^q e^{\theta_j x^{(i)}}} \quad (7)$$

### C. Overview of the Preliminary Classification Result

In this paper, both SAE and softmax layer adopt limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) as the optimization method that can be found in [26]. By the approach mentioned above, a single layer SAE with a softmax classified is trained. In this section, sigmoid function is chosen as the activation function and the parameters of hidden nodes number and sparsity are set to 100 and 0.2, suggested by Kunjin Chen in [24]. 10-fold cross validation (CV) is used to build up the model and the whole dataset consists of 7268 waveforms list in Table I. The preliminary overview of the accuracy of this classification framework is shown in Table III and IV.

TABLE III  
TRAINING CONFUSION MATRIX OF THE CLASSIFIER

	Air Gap	Float	Rod-plane	Surface	Accuracy
Air Gap	1557	65	42	37	91.53%
Float	20	1441	4	0	98.36%
Rod-plane	30	0	1389	8	97.34%
Surface	27	0	5	1974	98.40%

Average accuracy of 10-fold CV: 96.48%.

TABLE IV  
TESTING CONFUSION MATRIX OF THE CLASSIFIER

	Air Gap	Float	Rod-plane	Surface	Accuracy
Air Gap	172	5	3	4	93.48%
Float	0	148	0	0	100%
Rod-plane	6	0	164	0	96.47%
Surface	3	0	0	197	98.50%

Average accuracy of 10-fold CV: 96.19%.

One out of 10 training and testing confusion matrix of the proposed classifying framework are shown in Table III and IV. During the training process of SAE, the cost function, equation (4), decreases from several hundreds to 1.55 after hundreds of iterations. Testing accuracy being close to training accuracy means that the model is suitable without the problem of overfitting. Three of four kinds of PD waveforms, float, rod-plane and surface, could be identified with an accuracy around 98%. However, air gap defect is somehow difficult to be selected compared to other 3 defects. In the following section, by tuning the hyper-parameters of the SAE, extra effort is needed to improve the performance of identifying air gap PD current waveform.

## V. INFLUENCE OF HYPER-PARAMETERS

Several parameters in this network are tunable such as  $\lambda$ ,  $\beta$ ,  $\rho$ , number of hidden nodes, depth of the SAE and the activation function. According to the preliminary classification result in the last section, both  $\lambda$  and  $\beta$  in the cost function, weight of regularization and KL term, seem to be appropriate since the bias and variance of the classification result are both low enough. Thus, except that these 2 parameters are fixed to 0.003 and 5, others will be determined step by step. In this section, the training and testing process are completed by 10-fold cross-validation. 10 testing results are indicated by box plot with mean value being connected together.

### A. Sparsity Parameter

The sparsity parameter  $\rho$  controls the average activation level of all hidden nodes. For example, if  $\rho$  is set to 0.1, only 10% of the output range of the activation function is used. However, there is no specific guidance for selection of sparsity. The best sparsity should be a small value that can constrain the activation of hidden nodes to a low level. Yet it cannot be too small, for the extremely small sparsity will make the hidden nodes not be effectively activated. Typically, as introduced in [24], the optimized value of  $\rho$  is inside the range from 0.01 to 0.35.

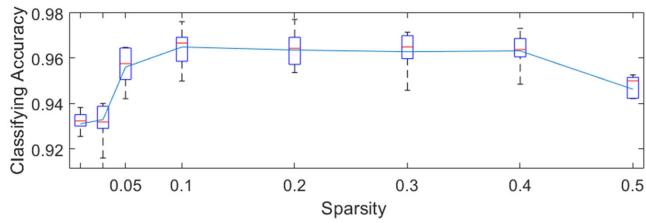


Fig. 8. Influence of sparsity parameters on the classifying accuracy.

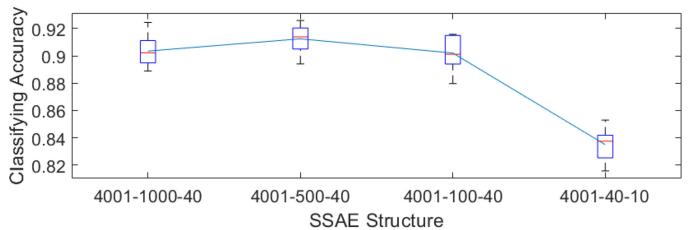


Fig. 10. Influence of SSAE structure on the classifying accuracy.

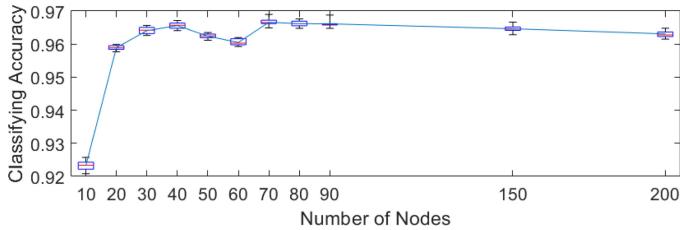


Fig. 9. Influence of number of hidden nodes on the classifying accuracy.

As shown in Fig. 8,  $\rho$  is set to 0.01, 0.03, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5. Since  $\rho$  is supposed to be a small value, the changing step of it is 0.02 below 0.1 and 0.1 above 0.1. When the sparsity is too small, for example 0.01 and 0.03, the accuracy of the classifier decreases to 92%-93%. As  $\rho$  increases, the testing accuracy is improved to an extend higher than 97%. However further increasing  $\rho$  from 0.4 to 0.5 will lead to a slight decline in the accuracy. Thus, sparsity is set to 0.1 in later training process.

#### B. Number of Hidden Nodes

Number of hidden nodes and depth of the network are the most important two parameters during the training of a deep learning model. No matter how many SAE layers are in the model, the size of the last SAE layer matters most because the features extracted from it are sent to softmax layer. Fig. 9 presents the series of cross-validation result of a single layer SAE using different number of hidden nodes with the sparsity of 0.1 whose activation function is sigmoid.

When the number of hidden nodes is too small, such as 10 in Fig. 9, the classifying accuracy drops dramatically because the AE cannot effectively extract the features representing the whole data set. Fig. 9 suggests that 4001-100 SAE structure has the highest accuracy above 97%. However, the variance of all the 10 cross-validation is also too high to be acceptable, so 40 and 70 both seem to be the desired number. Therefore, 40 is recommended as the final number of hidden nodes because of simplicity of the network.

#### C. Depth of the Framework

Once the size of the final layer has been determined, extra effort should be made to make sure whether it is necessary to use a deeper network. Cross-validation result of 4 different SSAEs is indicated in Fig. 10, from which conclusion could be drawn that deeper network seems not to be suitable for this scenario. Even though 4001-500-40 SSAE has the highest accuracy among the

three SSAE structures, it is still unmatched to the single layer SAE with the structure of 4001-40. The same result appears again during the comparison between 4001-40-10 SSAE and 4001-10 SAE. It can be concluded that a single layer SAE is enough for the feature extraction of PD waveforms and the original PD data will lose some information during the process of building feature through a single SAE. This information loss will be more severe as the SAE growing deeper.

#### D. Change of Activation Function

Activation function enables a deep learning model to approximate to a nonlinear model instead of a simple linear one, which makes the model more adaptive. In the former training process, sigmoid function is used as the default activation function, which is one of the most widely used activation functions today. Several concerns are aroused by the application of sigmoid and the most important one is gradient vanishment. In order to avoid this, data set should be preprocessed to an interval of (0, 1) to only make use of the steep part of the function.

Relu is another kind of function with growing popularity these days. It gives an output  $x$  if  $x$  is positive and 0 otherwise and this output characteristic will bring two advantages. For the input  $x > 0$ , the derivative of the function is a constant which will not cause vanishing gradient during the backpropagation and computation load is much less than sigmoid. For input  $x < 0$ , the output of the function is set to 0 and the sparsity of the network is therefore achieved directly, which means a simple AE instead of SAE will be enough under this condition.

Use  $z$ -score normalization method instead of linear normalization to [0, 1] to preprocess the data set and revise the cost function to equation (4) before Relu is applied as the activation function. Training process of the two functions is shown in Fig. 11. It clearly indicates that the training efficiency with Relu function is much higher the one with sigmoid. The running time of sigmoid function after 200 iterations is 475.74 seconds with the final normalized cost of 0.0073 while that of the Relu function is 424.16 seconds with the final normalized cost of 0.0066. Two confusion matrixes are summed up in Table V and VI.

Comparing with the result shown in Table III and IV, the later proposed structure, 4001-40 AE layer (with Relu function) combined with soft max layer, performs better. The overall classifying accuracy is up to 99.7%.

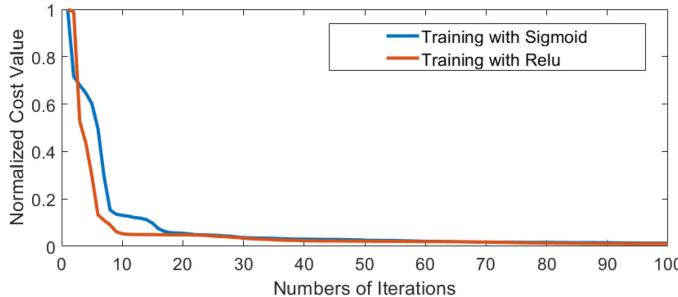


Fig. 11. Decreasing trend of cost value during different iteration in the training process with two kinds of activation function.

TABLE V  
TRAINING CONFUSION MATRIX OF THE CLASSIFIER

	Air Gap	Float	Rod-plane	Surface	Accuracy
Air Gap	1679	0	1	8	99.47%
Float	0	1439	7	0	99.52%
Rod-plane	0	0	1427	0	100%
Surface	1	10	0	1971	99.45%

Average accuracy of 10-fold CV: 99.56%.

TABLE VI  
TESTING CONFUSION MATRIX OF THE CLASSIFIER

	Air Gap	Float	Rod-plane	Surface	Accuracy
Air Gap	197	0	0	0	100%
Float	0	137	0	0	100%
Rod-plane	0	0	170	0	100%
Surface	1	2	0	221	98.66%

Average accuracy of 10-fold CV: 99.44%.

### E. Comparison With Other Traditional Machine Learning Method

After the determination of parameters discussed above, the final network consists of a 4001-40 AE layer with Relu function activating the hidden nodes and a softmax layer to deal with the classification job, the classifying accuracy is improved from 96.19% to 99.44%.

Feature extraction method applied on PD current waveforms used in previous research is dimensionality reduction such as PCA [27], [28] and T-SNE [29]. After feature extraction, traditional machine learning algorithm is used to cluster or classify the featured data such as *k*-means, support vector machine (SVM) and artificial neural network (ANN).

The classifying accuracy of different combinations of feature extraction and machine learning methods mentioned above are listed in Table VII. It is obvious that proposed method performs much better than traditional methods in this case. When using supervised classification method such as SVM and ANN, feature extraction does not seem to be very effective. On the other hand, when unsupervised algorithm is used, dimensionality reduction method of T-SNE can improve the accuracy a little, but it still makes no match to the supervised ones.

TABLE VII  
COMPARISON WITH OTHER TRADITIONAL METHODS

Method	Classifying Accuracy
4001-40 AE (with Relu) + Softmax	99.44%
4001-100 SAE + Softmax (Before parameter tuning)	96.61%
Raw Data + SVM[27]	74.07%
Raw Data + ANN[28]	77.37%
Raw Data + <i>k</i> -means[29]	26.57%
PCA [29-30]+ SVM	30.33%
PCA + ANN	70.13%
T-SNE[31] + SVM	30.83%
T-SNE + ANN	45.26%
T-SNE + <i>k</i> -means	39.38%

### VI. GENERAL VALIDITY TEST

Before this proposed method is applied to PD detection on real equipment, a more comprehensive blind test should be designed to validate the effectiveness.

#### A. Multiple Defect and Different Experiment Layout

1) *Change of the Defect Parameters and Topology:* The previous DL model was trained by four simple artificial defects. Though it has been stated by literatures [22]–[32] that four kinds of defect models described in Section II is representative to some real PD defects, the dimension parameters of real PD in the power equipment are various, the influence of the parameters on PD characteristics needs to be investigated. In this paper, thickness of the air gap and surface are altered, the distance between floating and main electrode as well as the rod and plane electrode are changed. All the parameters of different defect model are summarized in Table VIII.

2) *Alternative Test Circuit:* In the original test layout in Fig. 3,  $C_k$  is used as the coupling capacitor to couple the PD pulse and detected by the impedance  $Z_d$ . The FFT shape can be modified due to the dynamic characteristics of the transmission line interposed between the artificial PD source and coupling branch,  $C_k$  and  $Z_d$ . The influence of this branch must be considered. Besides, AC high-voltage source used in Fig. 3 is an ideal non-PD AC generator. It should be substituted by a real power transformer considering the practical operation.

The power transformer used in Fig. 11 is a 10 kVA 10kV/400V oil-filled transformer, the winding connection of which is Yy0. When the primary (low voltage) side is applied a 3-phase 380V AC voltage, the RMS value of the phase voltage,  $U_{A0}$ , of the secondary (high voltage) side should be calculated as

$$U_{A0} = \frac{380}{\sqrt{3}} * \frac{10000}{0.4} = 5.48 \text{ kV}$$

In this way, three test circuits could be used to acquire PD current waveforms. Original circuit described in Fig. 3 is marked as circuit 1. Remove  $C_k$  branch in circuit 1 and mark it as circuit 2. Test layout shown in Fig. 12 is circuit 3. In order to inspect the robustness of the DL model, all the PD defect models were split into 4 groups which are used in four different contrasting experiments respectively using different circuits. Detailed information could be referred to Table VIII.

**TABLE VIII**  
BRIEF SUMMARY OF ALL THE DEFECT MODELS

Purposes of the Data	Experiment Group (Test Circuit)	Model Type	Specific parameters (mm)	Applied voltages (kV)	Number of waveforms	Data tag
Training & Validating	1 (1)	Air Gap	2.7	9, 10, 11	1884 (750, 528, 606 respectively)	1
		Float	1.92	11.6, 13, 14	1581 (512, 527, 542 respectively)	2
		Rod-plane	8.1	6, 4.5, 5.6	1597 (555, 516, 526 respectively)	3
		Surface	5	9, 11, 12.5	2206 (427, 631, 1103 respectively)	4
Testing the influence of dimension parameters of PD	2 (1)	Air Gap	1.5	8, 9, 10	1820 (605, 599, 616 respectively)	5
		Float	1	9.4, 13, 15	1375 (166, 696, 513 respectively)	6
		Rod-plane	5	2.9, 3.8, 4.5	1753 (560, 608, 585 respectively)	7
		Surface	4	7, 9, 10.8	1643 (394, 685, 564 respectively)	8
Testing the influence of capacitor branch	3 (2)	Air Gap	4	9.8, 11.1, 12	1881 (615, 623, 643 respectively)	9
		Float	5	15	1345	10
		Rod-plane	10	4.7, 5.2, 5.8	1826 (559, 639, 628 respectively)	11
Testing in real transformer	4 (3)	Rod-plane	6	9.4, 11, 13	1505 (486, 474, 545 respectively)	12
		Rod-plane	8, 10	5.5	561 (228, 333 respectively)	13

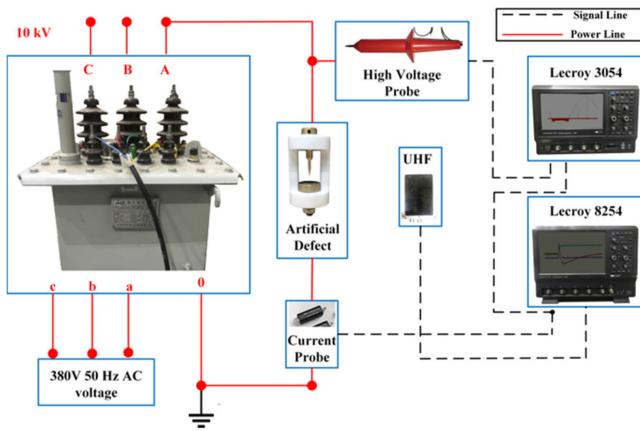


Fig. 12. Alternative test circuit.

As shown in Table VIII, data in group 1 are used as training and validation data set, with which DL model can be trained and improved by parameter tuning (already finished in the Section IV and V). The other 3 groups are going to be used as the testing data set to check the universality of the model. Since rod-plane PD waveform is the most common waveform during the previous research, so it is chosen as the example including in group 4.

### B. Result and Discussion

1) *Overall Testing Accuracy*: Refer to all the results of three blind tests list from Table IX to XI, the DL model trained in previous section is robust enough since all the 3 average accuracies exceed 96%.

**TABLE IX**  
TESTING CONFUSION MATRIX OF GROUP 2

	Air Gap	Float	Rod-plane	Surface	Accuracy
Air Gap	1719	101	0	0	94.45%
Float	0	1370	5	0	99.64%
Rod-plane	0	0	1753	0	100%
Surface	0	53	2	1588	96.65%

Average accuracy: 97.56%.

**TABLE X**  
TESTING CONFUSION MATRIX OF GROUP 3

	Air Gap	Float	Rod-plane	Surface	Accuracy
Air Gap	1867	0	14	0	99.26%
Float	2	1250	87	6	92.94%
Rod-plane	0	0	1810	16	99.12%
Surface	0	107	1	1397	92.82%

Average accuracy: 96.45%.

**TABLE XI**  
TESTING CONFUSION MATRIX OF GROUP 4

	Air Gap	Float	Rod-plane	Surface	Accuracy
Rod-plane	2	0	559	0	99.64%

2) *Discussion of High Frequency Disturbance*: Take the rod-plane PD current waveforms in all the four group as an example. FFT aggregated FFT plot of rod-plane data in each group are shown in Fig. 5(c) and Fig. 13(a) to (c). It is clear that at the frequency range below 10 MHz, frequency component of the PD waveform indicates small variation. It can be discovered that merely changing the parameter of a PD defect has no influence on the characteristic frequency (at 19 MHz, 95 MHz and 168 MHz). However, the dimension parameter does affect the

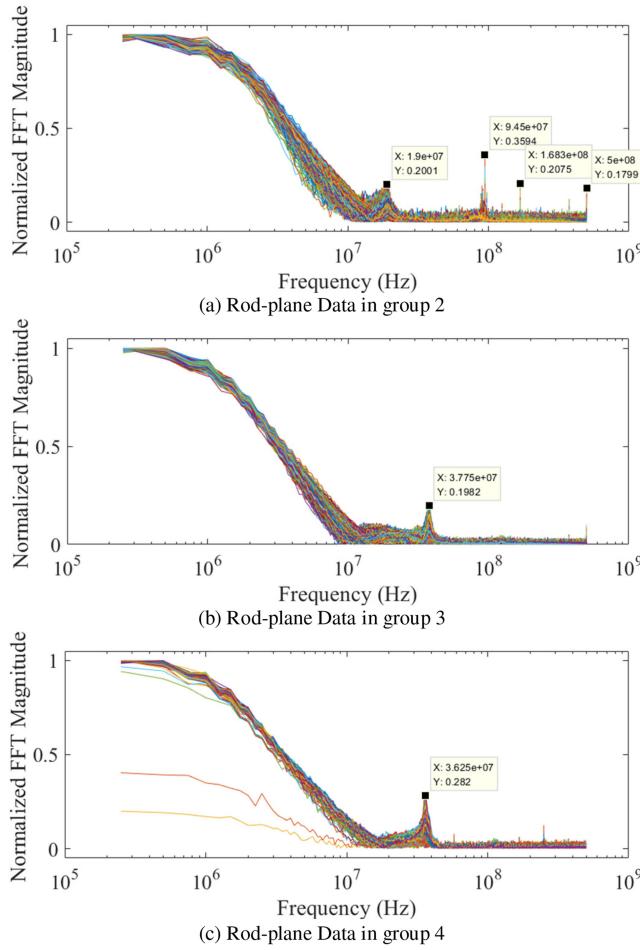


Fig. 13. Aggregate of FFT plot of all the acquired data.

magnitude of the high frequency component, for the magnitude of 95 MHz and 168 MHz in Fig. 13(a) are much lower than that in Fig. 5(c). After removing the coupling capacitor branch, characteristic frequencies at high frequency range disappear and meanwhile the first characteristic was moved from 19 MHz to 37 MHz. Besides, comparing with Fig. 13(c), there is a slight increase at 15 MHz in Fig. 13(b), which means that change of high voltage source does not have too much influence to the waveforms in frequency domain.

Based on the above analysis, the whole frequency domain could be divided into two parts, below and above 10 MHz. Frequency components below 10 MHz contain important information of PD waveform itself, while those above 10 MHz are disturbances brought by dynamic characteristics of the transmission line interposed between the artificial PD source and coupling branch, which can be neglected. From the testing result, it can be concluded that this novel method of PD detection and classification can be immune to high frequency disturbance.

As introduced in Section IV, final features used to train a softmax layer is extracted by AE that is trained by a large amount of waveforms. In this very case, the number of hidden nodes is 40, which determines the dimension of matrix  $W$  in the AE is

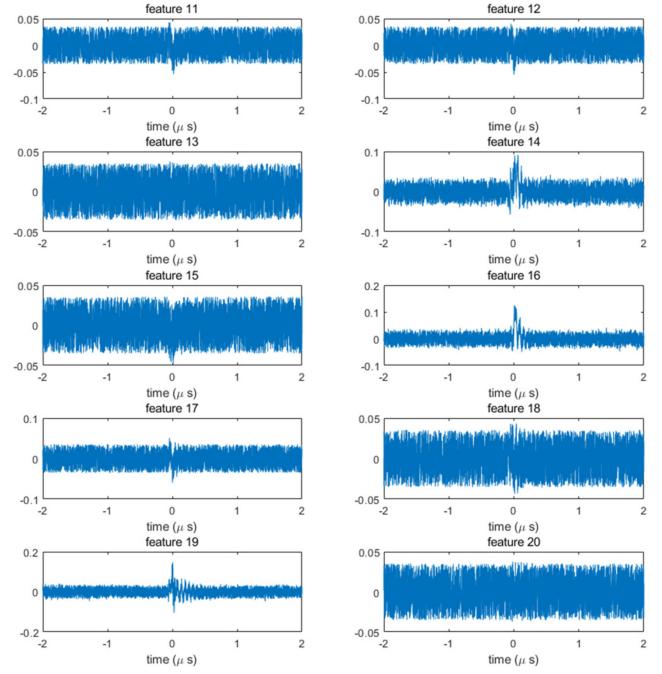


Fig. 14. Visualization of feature 11–20.

$40 \times 4001$ . Thus, the 40 features could be visualized by 40 vectors with 4001 input waveforms. Part of the features, 11–20, are shown in Fig. 14. Each feature waveform contains some of the information of a complete input PD signal. For example, feature 13 contains nothing of the pulse but background noise; feature 14 and 15 consist part of the single-direction pulse; feature 19 has the oscillation pulse components. When an input PD waveform is sent into AE, the encoding process realized by Eq. (1) will active some of the 40 features. In this way, if the input signal contains a new high-frequency disturbance that has never been learnt by AE, this disturbance will be automatically omitted by the system. This is the intrinsic nature of robustness of the proposed method.

### C. Cross Validation With Different Data in Different Groups

To further prove the general validity of the proposed method, a cross-validation among different models trained by different experiment groups, the result of which is listed in Table XII.

In Table XII, another 3 models are trained and validated. Model 2 and 3 are obtained by experiment group 2 and 3 respectively. Training and validating accuracy of both these 2 model are higher than 99.6%, which means that the models are trained successfully without over fitting. The overall testing accuracies of them with rest of the data is 97.69% and 96.65. Though they are not as high as the training accuracy, they are still at the same level that is much higher than that of the traditional method. Model 4 is learnt from the data selected from all the 3 experiment groups, the high classification accuracy of which further proves that the method proposed in this paper is robust enough.

TABLE XII  
CROSS VALIDATIONS RESULTS WITH DIFFERENT MODELS (%)

Experiment Group (Test Circuit)	Model Type	Data tag	Model 2			Model 3			Model 4		
			Training	Validation	Testing	Training	Validation	Testing	Training	Validation	Testing
1 (1)	Air Gap	1	-	-	97.93	-	-	92.82	-	-	97.56
	Float	2	-	-	99.87	-	-	95.57	100	100	
	Rod-plane	3	-	-	100	-	-	99.44	100	100	
2 (1)	Surface	4	-	-	98.96	-	-	98.64	-	-	95.42
	Air Gap	5	100	100	-	-	-	97.53	-	-	93.46
	Float	6	99.69	99.52	-	-	-	97.96	-	-	95.71
	Rod-plane	7	100	100	-	-	-	99.14	-	-	100
3 (2)	Surface	8	99.91	99.59	-	-	-	91.54	99.91	99.59	
	Air Gap	9	-	-	91.87	99.7	100	-	99.77	100	
	Float	10	-	-	92.86	99.89	99.00	-	-	-	92.42
4 (3)	Rod-plane	11	-	-	99.07	100	100	-	-	-	93.10
	Surface	12	-	-	99.21	99.05	99.33	-	-	-	98.41
Overall accuracy (%)			99.91	99.8	97.69	99.67	99.64	96.65	99.91	99.9	95.94

Model 2 is trained and validated by 70% and 30% data of tag 5,6, 7 and 8. Data with other tags is used as testing dataset.

Model 3 is trained and validated by 70% and 30% data of tag 9, 10, 11 and 12. Data with other tags is used as testing dataset.

Model 4 is trained and validated by 70% and 30% data of tag 9, 2, 3 and 8. Data with other tags is used as testing dataset.

## VII. CONCLUSION AND FUTURE WORK

This paper introduces a novel method to detect and identify different kinds of PD. PD current is used as the input signal other than traditional ERA and UHF signals. Four kinds of artificial PD defects were applied by AC voltage and thousands of current waveforms were acquired as the data set. Each kind of current waveform is validated by the detecting signal of ERA and UHF. Besides, phase relationship between applied AC voltage and current is compared with previous research.

Dimensionality reduction method such as PCA and T-SNE are used to squash the dataset to a 3-dimension vector to be easily visualized, the result of which shows the feasibility of classification of these four kinds PD waveforms. Proposed deep learning network of SAE and soft max shows a desirable classification result exceeding 96%. After a careful selection of parameters, sparsity is set to 0.1 and the size of a single layer SAE is designed as 4001-40. Moreover, change of activation function from sigmoid to relu could further improve the accuracy to 99.7%. Comparison to traditional identification method is also made to show the desirable performance of the proposed novel method.

A general validity test has been conducted. In this test, four kinds of models with different dimension parameters are manufactured; considering the high-frequency noise or disturbance in the test circuit, three kinds of test circuits including a real power transformer were applied during obtaining the data set. The testing accuracy is more than 96%. The result proves that both defect dimension parameters and high-frequency

disturbance have very little influence on the classifying, which means the proposed method has a strong robustness.

Conclusion in this paper indicates that PD current waveform could be directly used as the PD detection signal that can be successfully identified by deep learning algorithm, which means human-designed features are much more time-consuming and indirect. In the future, if this novel method is going to be perform under an on-site condition involves a lot of unknown noise or disturbance such as corona or other electrostatic disturbance, more on-site data should be collected to upgrade the model.

## REFERENCES

- [1] R. N. Wu and C. K. Chang, "The use of partial discharges as an online monitoring system for underground cable joints," *IEEE Trans. Power Del.*, vol. 26, no. 3, pp. 1585–1591, Jul. 2011.
- [2] *Partial Discharge Measurements*, IEC Standard 60270, 2000.
- [3] T. Hoshino, H. Koyama, S. Maruyama, and M. Hanai, "Comparison of sensitivity between UHF method and IEC 60270 for onsite calibration in various GIS," *IEEE Trans. Power Del.*, vol. 21, no. 4, pp. 1948–1953, Oct. 2006.
- [4] W. Gao, D. Ding, and W. Liu, "Research on the typical partial discharge using the UHF detection method for GIS," *IEEE Trans. Power Del.*, vol. 26, no. 4, pp. 2621–2629, Oct. 2011.
- [5] W. Gao, D. Ding, W. Liu, and X. Huang, "Investigation of the evaluation of the PD severity and verification of the sensitivity of partial-discharge detection using the UHF method in GIS," *IEEE Trans. Power Del.*, vol. 29, no. 1, pp. 38–47, Feb. 2014.
- [6] M. A. Brown *et al.*, "Correlation of partial discharge and dissolved gas analysis results from discharge activity in SRBP," in *Proc. IEEE Conf. Elect. Insul. Dielect. Phenomena*, 2007, pp. 268–271.
- [7] A. H. El-Hag *et al.*, "Online oil condition monitoring using a partial-discharge signal," *IEEE Trans. Power Del.*, vol. 26, no. 2, pp. 1288–1289, Apr. 2011.

- [8] X. Wang *et al.*, "Acousto-optical PD detection for transformers," *IEEE Trans. Power Del.*, vol. 21, no. 3, pp. 1068–1073, Jul. 2006.
- [9] H. Wang *et al.*, "Research of UHF calibration technique for four kinds of partial discharge defects in GIS," *Power Syst. Protection Control*, vol. 40, no. 21, pp. 7–13, 2012.
- [10] K., R. Kurrer and K. Feser, "The application of ultra-high-frequency partial discharge measurements to gas-insulated substations," *IEEE Trans. Power Del.*, vol. 13, no. 3, pp. 777–782, Jul. 1998.
- [11] G. P. Cleary and M. D Judd, "UHF and current pulse measurements of partial discharge activity in mineral oil," *IEE Proc. Sci., Meas. Technol.*, vol. 153, no. 2, pp. 47–54, Mar. 2006.
- [12] J. Tang *et al.*, "Assessment of PD severity in gas-insulated switchgear with an SSAE," *IET Sci., Meas. Technol.*, vol. 1, no. 4, pp. 423–430, 2017.
- [13] L. Hao and P. L. Lewin, "Partial discharge source discrimination using a support vector machine," *IEEE Trans. Dielect. Electr. Insul.*, vol. 17, no. 1, pp. 189–197, Feb. 2010.
- [14] I. J. Seo, J. Y. Koo, Y. J. Lee, B. W. Lee, J. T. Kim, and J. H. Lee, "Identification of insulation defects by modified chaotic analysis of partial discharge under DC stress," CIGRE Session, Paris, France, D1-310, pp. 1–8, 2012.
- [15] A. Pirker and U. Schichler, "Partial discharges at DC voltage-measurement and pattern recognition," in *Proc. Condition Monit. Diagnosis*, 2016, pp. 287–290.
- [16] M. S. Siregar *et al.*, "Partial discharge patterns on cross-linked polyethylene DC power cables," in *Proc. Power Eng. Renewable Energy*, 2016, pp. 119–123.
- [17] L. Satish and W. S. Zaengl, "Artificial neural networks for recognition of 3-d partial discharge patterns," *IEEE Trans. Dielect. Electr. Insul.*, vol. 1, no. 2, pp. 265–275, Apr. 1994.
- [18] S.M. Strachan and S. Rudd, "Knowledge-based diagnosis of partial discharges in power transformers," *IEEE Trans. Dielect. Electr. Insul.*, vol. 15, no. 1, pp. 259–268, Feb. 2008.
- [19] J. Li and R. Liao, "Oil-paper aging evaluation by fuzzy clustering and factor analysis to statistical parameters of partial discharge," *IEEE Trans. Dielect. Electr. Insul.*, vol. 17, no. 3, pp. 756–763, Jun. 2010.
- [20] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [21] 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.
- [22] K. Li *et al.*, "Analysis of air decomposition by-products under four kinds of partial discharge defects," *IEEE Trans. Dielect. Electr. Insul.*, vol. 24, no. 6, pp. 3713–3721, Dec. 2017.
- [23] A. Ng, "Sparse autoencoder," CS294A Lecture Notes, vol. 72, pp. 1–19, 2011.
- [24] K. Chen, J. Hu, and J. He, "A framework for automatically extracting overvoltage features based on sparse autoencoder," *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 594–604, Mar. 2018.
- [25] C. Do and A. Y. Ng, "Transfer learning for text classification," in *Advances in Neural Information Processing System*. Cambridge, MA, USA: MIT Press, 2006, pp. 299–306.
- [26] J. Ngiam *et al.*, "On optimization methods for deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2011, pp. 265–272.
- [27] P. L. Lewin, L. A. Petrov, and L. Hao, "A feature based method for partial discharge source classification," in *Proc. IEEE Int. Symp. Elect. Insul.*, 2012, pp. 443–448.
- [28] R. M. Sharkawy *et al.*, "SVM classification of contaminating particles in liquid dielectrics using higher order statistics of electrical and acoustic PD measurements," *IEEE Trans. Dielect. Electr. Insul.*, vol. 14, no. 3, pp. 669–678, Jun. 2007.
- [29] D. Thummapal, A. Jain, and B. E Kushare, "Partial discharge pattern recognition of HV GIS by using artificial neural networks," *Int. J. Eng. Res. Technol.*, vol. 3, no. 11, pp. 449–454, 2014.
- [30] R. Liao, Y. Fernandes, K. Tavernier, G. A. Taylor, and M. R. Irving, "Recognition of partial discharge patterns," in *Proc. Power Energy Soc. Gen. Meeting*, 2012, pp. 1–8.
- [31] V. P. Darabad *et al.*, "Data mining on partial discharge signals of power transformer's defect models," *Int. Trans. Elect. Energy Syst.*, vol. 23, no. 3, pp. 423–437, 2013.
- [32] 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, 2002.

**Lian Duan** was born in Zhengzhou, China, 1989. He received the B.Sc. and M.Sc. degrees in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, in 2012 and 2015. He is currently working toward the Ph.D. degree with the Department of Electrical Engineering, Tsinghua University, Beijing, China. His research interests include deep learning and data mining in fault diagnosis of power equipment.

**Jun Hu** was born in Ningbo, China, in 1976. He received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Department of Electrical Engineering, Tsinghua University, Beijing, China, in 1998, 2000, and 2008, respectively. He is currently an Associate Professor with the Department of Electrical Engineering, Tsinghua University. His research fields include overvoltage analysis in power system, sensors and big data, dielectric materials, and surge arrester technology.

**Gen Zhao** received the B. S. degree from Chongqing University, Chongqing, China, and the Ph.D. degree from Tsinghua University, Beijing, China, both in electrical engineering, in 2013 and 2018, respectively.

He is currently a Postdoctoral Researcher with the Department of Electrical Engineering, Tsinghua University, Beijing, China. His research interests include inverse problem and data mining in energy systems.

**Kunjin Chen** was born in Changsha, China, in 1993. He received the B.Sc. degree in electrical engineering from Tsinghua University, Beijing, China, in 2015, where he is currently working toward the Ph.D. degree with the Department of Electrical Engineering. His research interests include pattern recognition and data mining in power systems.

**Jinliang He** (M'02–SM'02–F'08) was born in Changsha, China, in 1966. He received the B.Sc. degree in electrical engineering from the Wuhan University of Hydraulic and Electrical Engineering, Wuhan, China, in 1988, the M.Sc. degree in electrical engineering from Chongqing University, Chongqing, China, in 1991, and the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 1994.

He was a Lecturer and an Associate Professor in 1994 and 1996, respectively, with the Department of Electrical Engineering, Tsinghua University. From 1997 to 1998, he was a Visiting Scientist with Korea Electrotechnology Research Institute, Changwon, South Korea, involved in research on metal oxide varistors and high voltage polymeric metal oxide surge arresters. From 2014 to 2015, he was a Visiting Professor with the Department of Electrical Engineering, Stanford University, Palo Alto, CA, USA. In 2001, he was promoted to a Professor with Tsinghua University. He is currently the Chair with High Voltage Research Institute, Tsinghua University. He has authored five books and 400 technical papers. His research interests include overvoltages and EMC in power systems and electronic systems, lightning protection, grounding technology, power apparatus, and dielectric material.

**Shan X. Wang** (M'88–SM'06–F'09) received the B.S. degree in physics from the University of Science and Technology of China, Hefei, China, in 1986, the M.S. degree in physics from Iowa State University, Ames, IA, USA, in 1988, and the Ph.D. in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, PA, USA, in 1993.

He is currently the Director with the Stanford Center for Magnetic Nanotechnology and a Professor of Materials Science and Engineering, jointly of Electrical Engineering, and by courtesy, a Professor of Radiology with Stanford University, Stanford, CA, USA.