A Convolutional Neural Network Based Deep Learning Methodology for Recognition of Partial Discharge Patterns from High Voltage Cables
Xiaosheng Peng, Fan Yang, Ganjun Wang, Yijiang Wu, Li Lee, Zhaohui Li, Ashfaque Ahmed Bhatti, Chengke Zhou, Donald M. Hepburn, Alistair J. Reid, Martin D. Judd, W. H. Siew

Abstract—It is a great challenge to differentiate Partial Discharge (PD) induced by different types of insulation defects in high voltage cables. Some types of PD signals have very similar characteristics and are specifically difficult to be differentiate, even for the most experienced specialists. To overcome the challenge, a Convolutional Neural Network (CNN) based deep learning methodology for PD pattern recognition is presented in this paper. Firstly, PD testing for five types of artificial defects in Ethylene-Propylene-Rubber (EPR) cables was carried out in the High Voltage (HV) laboratory to generate signals containing PD data. Secondly, 3500 sets of PD transient pulses were extracted and then 33 kinds of PD features established. The third stage applies a CNN to the data: typical CNN architecture and the key factors which affect the CNN based pattern recognition accuracy are described. Factors discussed include the number of the network layers, convolutional kernel size, activation function, and pooling method. The paper presents a flowchart of the CNN based PD pattern recognition method and an evaluation with 3500 sets of PD samples. Finally, the CNN based pattern recognition results are shown and the proposed method is compared with two more traditional analysis methods, i.e. Support Vector Machine (SVM) and Back Propagation Neural Network (BPNN). The results show that the proposed CNN method has higher pattern recognition accuracy than SVM and BPNN, and that the novel method is especially effective for PD type recognition in cases of signals of high similarity, which is applicable for industrial applications [6-9].

Index Terms—Convolutional neural network, deep learning, high voltage cables, partial discharge, pattern recognition.

I. INTRODUCTION

PARTIAL Discharge (PD) monitoring plays an important role in evaluating the condition of cable insulation and in improving the reliability of power transmission and distribution systems [1-5]. On-line or off-line PD based condition monitoring and fault diagnosis have been increasingly reported in industrial applications [6-7]. However, differentiating defect types through PD pattern recognition is one of the most difficult challenges and has restricted the large-scale industrial application of PD based condition monitoring. PD signals induced by insulation defects which show a high similarity in the PD pattern structure are especially difficult to distinguish [6-9].

To overcome the challenge of differentiating defect type, different pattern recognition methods have been applied to detect PD patterns. These methods include Decision Tree (DT), Back Propagation Neural Network (BPNN), Support Vector Machine (SVM), and Rough Set (RS) Theory [10-12]. SVM and BPNN have been most widely used due to their excellent and stable pattern recognition performance [10-12]. Hao et. al. [10] combined Wavelet Transform (WT) and SVM to identify PD signals from different sources. Tang et. al. [11] employed Particle Swarm Optimization (PSO) to optimize the parameters of SVM applied to distinguish PD signals from different sources in Gas Insulated Switchgear (GIS). Li et. al. [12] combined, two types of BPNN, with the input parameters of Phase Resolved Partial Discharge (PRPD) pattern and Time-Resolved Partial Discharge (TRPD) pattern to improve the pattern recognition accuracy of different types of PD signals from GIS.

However, it has been recognized that the traditional machine learning methods, such as SVM and BPNN, have met bottlenecks in their development, which restrict further improvements in pattern recognition accuracy. In recent years, deep learning has been important in the development of Artificial Intelligence (AI) and pattern recognition systems [13-15]. Deep learning methods use deep neural networks employing multiple nonlinear layers. These have the ability to capture high dimensional nonlinearity and complex correlation which cannot be learned by traditional shallow neural network structures [16]. Deep learning methods are capable of self-learning and feature abstracting from massive datasets, which brings new perspectives and opportunities for PD pattern recognition.

Common deep learning methods include Convolutional Neural Network (CNN) [15], Stacked Denoising Auto-Encoders (SDAE), Recurrent Neural Network (RNN) and Deep Belief Network (DBN). CNN based deep learning algorithms have been successfully applied in many fields, such as speech recognition and image recognition, and are considered to be of increasing importance in the power industry [13-15]. In comparison to other deep learning methods, the complexity of CNN and the difficulty of training them are
greatly reduced through the parameters of sharing and local connection, which also reduce the risk of over-fitting [16]. According to [16], it is relatively easy to construct a network with deep architecture using CNN.

A CNN based deep learning methodology for PD pattern recognition of high voltage cables is presented in this paper and evaluated with 3500 sets of data. The data was generated using 5 types of PD fault under laboratory conditions. The proposed method is compared with two traditional machine learning methods, SVM and BPNN. The results show that CNN based PD pattern recognition is more effective than SVM and BPNN and is especially effective at differentiating PD types where the patterns have high similarity. This method is applicable not only to PD pattern recognition in HV cables but also to analysis of signals from other HV apparatus, such as gas-insulated substations, transformers and generators.

II. PD DATA GENERATION AND FEATURE EXTRACTION

A. EPR Cable and Artificial Defects

Five types of insulation defects found in EPR cables were simulated using a section of cable which has been in service. The artificial faults were stressed in a high voltage laboratory. The EPR cable comprises of an aluminum core, an inner semiconductor layer, an EPR layer, an outer semiconductor layer, a shielded copper strip, an aluminum armor layer and a polyvinyl chloride (PVC) sheath. A schematic of the EPR cable structure and a schematic of the defect structures are shown in Fig. 1.

![Diagram of EPR cable structure and artificial defects](image)

Fig. 1. The cable structure and the artificial defects.

Defect type 1 simulates a void crossing from the PVC oversheath to the EPR insulation layer. The defect was created through the following procedure: a cylindrical cavity was generated by a precision 0.4 mm diameter Printed Circuit Board (PCB) drill, and a copper strip in contact with the outer semiconductor fixed on the circular hole to seal the defect.

Defect type 2, simulating a protrusion on the outer conductor, was created by inserting a 0.4 mm diameter PCB drill into the cable. The drill remains in touch with the conductive parts.

Defect type 3, simulating a floating protrusion on the outer conductor was created by inserting a PCB drill into the cable but ensuring that the conductive parts do not touch the drill.

Defect type 4, simulating a cable having external damage, introduced a breach in the outer semiconductor of the cable. To replicate this fault type, an area 7mm × 7mm was cut from the PVC sheath, the semiconductor layer and the shielding layer.

Defect type 5 simulates PD from an end termination. It was generated by connecting a part of the outer copper sleeve of the cable to the system earth.

Given the similarity of construction of defects type 2 and type 3, PD signals generated by these defects might be a challenge to separate.

B. Experimental Setup and PD Pulse Feature Extraction

A commercial High Frequency Current Transformer (HFCT), having a bandwidth from 20 kHz to 20 MHz, as shown in Fig. 2, was used for PD signal extraction.

The principle of the PD testing system, which is based on the IEC60270 system, is shown in Fig. 3. EPR cable samples of lengths 1-2 meters, with the artificial defects described in section 1.1, were connected to the high voltage supply for PD testing. The HFCT was placed around the earth strap of the EPR cable to capture PD signals. A wide-band amplifier was connected to the HFCT to adjust the signal from HFCT to the input limits of the data acquisition (DAQ) system, which in this case was a LeCroy 104 Xi oscilloscope. A voltage probe was connected to the high voltage source to acquire the 50 Hz phase reference for future PD pattern analysis. The DAQ rate of the oscilloscope was set to 100MS/s and the DAQ time length was set to 20ms (one power cycle of the 50Hz supply).

![Diagram of experimental setup for PD measurement](image)

Fig. 2. HFCT frequency response.

The test voltage was increased from 0 to PD Inception Voltage (PDIV), in steps of 1kV, up to the maximum test voltage. For defect types 1, 4 and 5, the maximum value was 11kV, the rated voltage of the EPR cable. For defect types 2 and 3, the maximum voltage was 13 kV and 12 kV respectively. Raw data was logged for each test. The number of sets of raw PD data for each test voltage of the five types of defects is shown in Table 1.

From the raw data, PD transient pulses were extracted to generate PD data sets for five types of artificial defect. From the complete data 700 transient pulses were randomly selected from the data for each defect type for further analysis. This analysis relied upon feature extraction, as discussed in the
C. Feature Extraction

Based on the 3500 samples of PD transient pulses, feature extraction was carried out in which important parameters of the signals were acquired. The PD pulse signals were considered in terms of the following 17 parameters: pulse width [17], rise time [17], fall time [17], peak voltage, pulse polarity, mean voltage, root mean square (RMS) voltage, standard deviation of peak pulse voltage, skewness of peak pulse voltage, kurtosis of peak pulse voltage, voltage crest factor (peak magnitude over RMS), voltage form factor (RMS over average value), main frequency of transient pulse, phase angle, equivalent time length (T) [17], equivalent bandwidth (W) [17] and discharge magnitude in picocoulombs.

In addition to the 17 PD features described above, Wavelet Transform (WT) analysis was applied to the transient pulses [10]. The input signals were decomposed into five layers and 16 wavelet features were established for the wavelet transform. Finally, 33 types of PD features were constructed and were applied as the input parameters of the pattern recognition methods, BPNN, SVM, and CNN.

The principle of decomposition of PD to extract features of transient PD pulses is shown in Fig. 4.

![PD feature construction of transient PD pulses based on the wavelet transform.](image)

In Fig. 4, the original transient PD pulses were decomposed into 5 layers based on a db5 mother wavelet. The wavelet features of the signals were constructed according to the wavelet coefficients. $A_i$ represents the approximate composition of each layer. $x$ represents the original signal. $d_i$ represents the detail component of each layer.

16 wavelet features are named as $Ed_1\sim Ed_5$, $Ea_1\sim Ea_5$, $ED_1\sim ED_5$ and $EA_5$ respectively.

$Ed_i$ represents the detailed energy of layer $i$, the formula for which is shown in equation (1).

$$Ed_i = \int d_i^2(t) dt$$ (1)

$Ea_i$ represents the global energy of layer $i$, the formula for which is shown in equation (2).

$$Ea_i = \int A_i^2(t) dt$$ (2)

$ED_i$ represents the ratio of detail energy of layer $i$ and the energy of the original signal, the formula for which is shown in equation (3).

$$ED_i = \frac{\int d_i^2(t) dt}{\int x^2(t) dt}$$ (3)

$EA_5$ represents the ratio of global energy of the fifth layer and the energy of original signal $x$, the formula for which is shown in equation (4).

$$EA_5 = \frac{\int A_5^2(t) dt}{\int x^2(t) dt}$$ (4)

D. Initial Analysis

An initial comparison of the data from the five fault types was carried out. Fig. 5 shows one set of PD raw data, with a time length of 20 ms, for each defect type, and one typical fast-transient PD pulse within said data plotted on a shorter time-base to allow additional detail of the PD signals to be seen. Phase-Resolved PD Patterns were constructed using two different PD parameters, namely discharge magnitude and equivalent time length. The patterns generated for each of the parameters is shown in Fig. 6.

![One set of PD raw data and one typical PD transient pulse for five types of cable defects.](image)

From Fig. 6, the following conclusions can be drawn: Firstly, PD type 2, in red, and PD type 3, in blue, are very similar, as the two PRPD patterns of these faults overlap. This confirms the comment in section II.A. Secondly, PD type 4 and PD type 5 show differences to PD types 1, 2 and 3. For example, in Fig. 5(a), the PRPD of discharge magnitude, data points for type 4 and 5 have a lower degree of overlap with PD types 1, 2 and 3.
From Fig. 5 and Fig. 6 it can be seen that the 3500 sets of samples from 5 fault types were ideal data for pattern recognition method evaluation, as 3 types of PD signal had high mutual-similarity but could be distinguished from the other defect types.

Additional analysis would be required to separate the fault types. The Convolutional Neural Network method was selected, as outlined in the following section.

![Convolutional Neural Network](image)

**III. CONVOLUTIONAL NEURAL NETWORK**

CNN evolved from the structure of the biological vision system, which has the ability to take specific features and figure out patterns in visual images [15, 18]. In the 1960s, Hubel et al. discovered that visual information transmitted from the retina to the brain was processed by stimulating respective fields with multiple levels [18]. In 1980, Fukushima first proposed the Neocognitron [18], which is a self-organizing multilayer neural network. In 1998, LeCun et al. [19] proposed the first type of convolutional neural network, LeNet-5, by utilizing a gradient-based back propagation algorithm for supervised training. In 2012, Krizhevsky et al. won the image classification competition, ImageNet Large Scale Visual Recognition and Challenge, with AlexNet [20-21]. CNN is widely applied in image recognition and speech recognition because of the excellent feature learning ability.

**A. Typical Architecture of CNN**

The typical structure of CNN is shown in Fig. 7. As shown in the lower layer labels, the CNN structure is comprised of one or more alternately connected convolutional layers, pooling layers and a fully connected layer for classification. In general, the input to a CNN is a 1 or 2 dimensional (1 or 2D) data matrix.

The convolutional layer is utilized directly to extract features from the 1D or 2D matrixes and map the extracted features to form new feature maps. The structure in which multiple convolutional layers are distributed with pooling layers allows the ability to reduce the size of feature maps and gradually set up features with a high degree of spatial and uniform structure. The fully connected layer is applied to estimate the PD pattern, based on the extracted features sequentially conveyed by multiple convolutional and pooling layers. In most cases, CNN based pattern recognition systems are trained by a gradient descent algorithm [18]. Fundamentally, CNN is a mathematical model which processes raw input through data transformation and dimensionality reduction with multiple levels to a new feature representation field [18].

![CNN Structure](image)

There are multiple key factors which affect the pattern recognition accuracy of CNN, including the number of network layers, the size of the convolutional kernel, the type of activation function and the pooling method, as discussed in the following section.

**B. The Number of Network Layers**

CNN’s have a special structure in which the network nodes of the adjacent two layers are locally connected. In addition, CNN systems use shared parameters of connection weights in the same convolutional layers or pooling layers [18-20]. These characteristics greatly decrease the number of parameters, reduce the complexity of network training and the risk of over fitting [18-20]. Therefore, CNN has major advantages in constructing deep architecture neural networks and provides superior feature learning capability [19].

As the number of network layers increases, the feature abstraction and capabilities of CNN increase gradually but the number of parameters required for training also increases and the need for sample data volume rises accordingly, which will raise the risk of over fitting for small sample data [21-22]. However, there is no theoretical framework on how to determine the number of network layers [18-20]. Hence, for the successful application of CNN based pattern recognition, different numbers of network layers should be evaluated with the training data. The network with the highest pattern recognition accuracy should be chosen as the ideal CNN architecture.

**C. The Size of Convolutional Kernel**

The main block of the CNN convolutional layer has a series of convolutional kernels. During the forward propagation process, each convolutional kernel convolves with the input feature structure from the earlier layer to produce convoluted feature maps. The convoluted data is then subjected to a nonlinear activation function to get the output feature maps of the layer.

An example of 2D convolution is shown in Fig. 8. A 2×2 convolutional kernel convolves with the 3×3 input feature diagram to produce an output feature block with 4 units arranged in 2×2
grid. The convolutional kernel can be regarded as a feature detector then all the units in a feature map detect the same pattern but at different locations in the input feature layout.

The size of the convolutional kernel affects the performance of feature detecting directly. Moreover, the material complexity of each convolutional operation in the CNN model is theoretically linear with the side length of the convolutional kernel. If the size of convolutional kernel is too small, then the CNN model is susceptible to information loss in the process of feature learning, which will negatively affect the pattern recognition performance. If the size of the convolutional kernel is too big, the computational cost rapidly increases [19].

![Convolution and Feature Map](image)

**Fig. 8.** Sample of convolution: a 2×2 kernel convolves with a 3×3 input feature map to produce a 2×2 feature map.

### D. Nonlinear Activation Function

An activation function is used to introduce nonlinearities to the CNN, which plays an important role in determining the pattern recognition performance. Typical nonlinear activation functions contain hyperbolic tangent function (tanh), sigmoid, and rectified linear unit (relu) [21].

Google Brain proposed a new type of self-gated activation function, “swish”, which has an upper bound but no lower bound, accompanied with smooth and non-monotonic characteristics [21]. [21] pointed out that deep learning models with swish activation functions have better performance than the widely used relu activation function and that better results would be achieved by deep neural networks using swish if slightly smaller learning rates were adopted compared with neural networks with relu [21].

As the saturation property of sigmoid functions have the challenges of gradient disappearance and high computational cost, it is difficult to effectively train the limits of deep neural networks with sigmoids [18-21]. The tanh function has a similar saturation characteristic to sigmoids, giving rise to similar pattern recognition accuracy issues. In comparison to sigmoid and tanh, the linear and unsaturated characteristics of the relu function is favorable for making neural networks converge rapidly. However, some of the weights of nodes of the neural networks with relu may not be upgraded during training, i.e. a “dead zone” occurs. The absence of an upper bound and the smooth and non-monotonic characteristics are the advantages of the swish function over other existing activation functions.

### E. Pooling Method

As the original input feature block has a large number of features, the computational cost may become excessive if all the features from one convolutional layer are directly adopted. The pooling operation system is applied to collect semantically similar features in the pooling window within a feature structure into one feature. This helps to reduce the size of feature layout and achieve the goals of data dimensionality reduction [18-21].

There is some difference in the effectiveness of CNNs with different pooling approaches on various data sets. Currently, max pooling and average pooling are the two mainstream pooling methods [18-21].

For the convolved feature, the max pooling method selects the feature with the largest value within the pooling window to form the output feature. An example of the max pooling method is shown in Fig. 9. A 2×2 pooling window gets input from a convolved feature diagram and takes the feature with the greatest value in each window to generate a pooled feature. Max pooling discards non-maximal features which may result in information loss.

The average pooling method takes the average of the input features within the pooling window. Average pooling selects the average value of the features in the pooling window but this discrimination of data may also result in information loss.

The Rank based Average Pooling (RAP) method effectively handles the problem of the loss of useful information caused by average pooling and max pooling by weighted averaging of features in the pooling window [23]. RAP combines the advantages of max pooling and average pooling and therefore has better performance as it retains the discriminatory features lost in average pooling and the non-maximal features absolutely discarded in max pooling [23].

### IV. CNN BASED PD PATTERN RECOGNITION PROCESS

The flowchart of the CNN based PD pattern recognition method for high voltage cables is shown in Fig. 10.
The results of PD pattern recognition of CNN for different number of network layers were tested. The structure of each CNN had similar characteristics, i.e. the same number of convolutional and pooling layers and a single fully connected layer, for classification. In addition, the RAP pooling method, the swish activation function and convolutional kernel size of 1×4 were adopted. The first CNN had 3 layers (1 convolutional layer, 1 pooling layer and 1 fully connected layer), the second had 5 layers (2 convolutional, 2 pooling and 1 fully connected layer), up to the final CNN of 17 layers (8 convolutional, 8 pooling and 1 fully connected layer): results are shown in Fig. 11.

As the number of network layers increases, the ability of CNN to fit the reference data will increase, but the risk of over-fitting also increases, especially when the sample data is small. As there are 3,500 samples in the data set, which is a relatively small number, the optimal number of network layers should be relatively small. As shown in Fig. 11, among the 8 models generated, the 7-layer CNN has the highest overall PD pattern recognition accuracy and the highest pattern recognition accuracy for every type of defect.

The performance of CNN with different number of network layers was tested. The structure of each CNN had similar

![Flowchart of PD Pattern Recognition based on CNN.](image)

**V. RESULTS AND EVALUATION**

In this section, the PD pattern recognition results of CNN for high voltage cables were evaluated, in terms of number of network layers, the convolutional kernel size, the activation function and the pooling method. Furthermore, the proposed approach was compared with the traditional methods, i.e. BPNN and SVM.

The computer configuration was as follows: the operating system is Windows 7, the CPU is Intel i5-4200M and the memory size is 12 GB.

**A. Results of PD Pattern Recognition of CNN with Different Number of Network Layers**

The performance of CNN with different number of network layers was tested. The structure of each CNN had similar
importance to the PD pattern recognition performance of CNN. A small convolutional kernel size will restrict the feature abstracting and expressing capabilities of CNN, while a big convolution kernel size will cause redundant computation. From Fig. 12, it can be seen that the PD pattern recognition accuracy of CNN with a convolutional kernel size of 1×1, 1×4 and 1×8 is 82.10%, 92.57% and 92.57%, respectively. Furthermore, the training time of CNN with a convolutional kernel size of 1×1, 1×4 and 1×8 was 352s, 384s and 426s, respectively. As the size of the convolutional kernel was increased from 1×1 to 1×8, the PD pattern recognition accuracy of CNN increases at first and then plateaus. However, the training time has an approximately linear increase. From this, a CNN with convolutional kernel size of 1×4 has the highest pattern recognition accuracy and moderate computational cost, which means the best comprehensive pattern recognition performance. The results validate the analysis of the effect of convolutional kernel size on the pattern recognition performance of CNN in section III.C.

C. Results of PD Pattern Recognition of CNN with Different Activation Functions

The type of activation function applied strongly affects the effectiveness of CNN pattern recognition [21]. In this paper, the pattern recognition performance of CNNs with four activation functions, namely tanh, sigmoid, relu and swish, was compared, using a convolutional kernel size of 1×4 but retaining the other parameters as outlined in section V.B. The results are shown in Fig. 13.

From Fig. 13, the swish function has a higher pattern recognition accuracy than CNNs with the other three activation functions.

D. Results of PD Pattern Recognition of CNN Using Different Pooling Methods

In this section, the PD pattern recognition performance of CNNs with three pooling methods – average pooling, max pooling and RAP – are compared. The activation function adopted is swish, other parameters are the same as those in section V.C. The results are shown in Fig. 14.

From Fig. 14, CNN with RAP has the highest PD pattern recognition accuracy for every type of PD fault type. As discussed, RAP has the ability to merge the advantages of average pooling and max pooling, resulting in CNN with RAP having the best PD pattern recognition performance. The overall pattern recognition accuracy of CNN with RAP is 92.57%, which is 0.57% and 0.76% better than CNNs adopting max pooling and average pooling, respectively.

E. Pattern Recognition Results and Visualization of CNN

Based on the foregoing analysis, the final adopted architecture of CNN contains 3 convolutional layers, 3 pooling layers and 1 fully connected layer for classification. The other parameter configurations of CNN in this work are RAP pooling, swish activation function and convolutional kernel size of 1×4. The confusion matrix for PD pattern recognition of CNN is shown in Fig. 15.

As can be seen in Fig. 15, high error rates exist for defect types with high similarity, namely PD type 2 and type 3. The degree of confusion among the PD pattern recognition results of defect types 2 and 3 are: 21 of the 105 samples of defect type 3 are misidentified as type 2; 12 of the 105 samples of defect type 2 were misidentified as type 3. The result verifies the analysis of PRPD in section 1.3 and is the main reason for the need to improve overall PD pattern recognition accuracy.

In order to illustrate the feature abstracting and expressing capabilities of CNN, the t-SNE algorithm was applied to map the original PD features and the output of each convolutional layer of CNN to a two-dimensional plane for visualization [24]. The results are shown in Fig. 16. The t-SNE algorithm is a non-linear dimensionality reduction technique: further information on t-SNE can be found in [24].

In Fig. 16 it can be observed that data points for each defect type are grouped together in the original feature space. As expected, the degree of overlap for points corresponding to defect types 2 and 3 are particularly large.
The visualization results of each convolutional layer output in CNN are also shown in Fig. 16. From this it can be clearly observed that as the number of convolutional layers increases, the samples matching to the five defect types become systematically more separated. After being processed by the first convolutional layer, the clusters of five fault types have a greater degree of separation than those in the original feature space: the discrimination of types 1, 4 and 5 is relatively large. After being processed by the second convolutional layer, the relative separation of the five types of defect is greater than those in the previous convolutional layer. After being handled by the third convolutional layer, the sample point clusters are more concentrated and differentiable from each other. The contribution of each convolutional layer of CNN to distinguish the complementary samples of the five types of defect is visually indicated in Fig. 16, representing the principle of the layer-wise optimization of the deep neural network.

F. Comparison with the Traditional Pattern Recognition Methods

To express the advantages of PD pattern recognition based on the CNN method, PD pattern recognition by BPNN and SVM was also studied. Comparative results are shown in Table 2.

<table>
<thead>
<tr>
<th>Defect type</th>
<th>Type 1</th>
<th>Type 2</th>
<th>Type 3</th>
<th>Type 4</th>
<th>Type 5</th>
<th>Accuracy</th>
<th>Training Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>98.08%</td>
<td>80.87%</td>
<td>85.42%</td>
<td>99.06%</td>
<td>100.00%</td>
<td>92.57%</td>
<td>384.0s</td>
</tr>
<tr>
<td>BPNN</td>
<td>90.99%</td>
<td>70.87%</td>
<td>69.31%</td>
<td>99.05%</td>
<td>99.05%</td>
<td>86.10%</td>
<td>6.3s</td>
</tr>
<tr>
<td>SVM</td>
<td>92.59%</td>
<td>75.00%</td>
<td>72.82%</td>
<td>99.05%</td>
<td>99.05%</td>
<td>87.81%</td>
<td>403.3s</td>
</tr>
</tbody>
</table>

From Table 2, overall PD pattern recognition accuracy rate of CNN is 92.57%, which was the highest among the three methods. The overall accuracy of SVM is 87.81% and that of BPNN was 86.10%. BPNN was lowest among the approaches. Compared to BPNN and SVM, overall pattern recognition accuracy of CNN was increased by 6.47% and 4.76%, respectively. Furthermore, the pattern recognition accuracy of CNN for every type of defect was the highest among all methods. The degree of distinction of defect types 1, 4, and 5 is relatively high, making it easier to differentiate the three defect types. PD pattern recognition accuracy rates of CNN for fault types 1, 4 and 5 are 98.08%, 99.06%, and 100.00%, respectively. As discussed, defect types 2 and 3 have high similarity. CNN has a PD pattern recognition accuracy for defect type 2 of 80.87%, which was 10.00% and 5.87% better than BPNN and SVM. The accuracy of CNN for defect type 3 is 85.42%, which was 16.11% and 12.60%, better than BPNN and SVM respectively.

One disadvantage of CNN is that the method takes more training time than BPNN. The training time of CNN is 384.0 seconds and the training time of BPNN is 6.3 seconds. The training time of SVM is 403.3 seconds, which is close with CNN. The testing time for different methods, CNN, BPNN and SVM, is less than 1 second.

VI. CONCLUSIONS AND DISCUSSIONS

A convolutional neural network based deep learning methods for PD pattern recognition of different types of insulation defects in HV cables has been presented. The proposed technique was compared with traditional SVM and BPNN methods. The conclusions drawn were as follows:

- The CNN based PD pattern recognition method proposed is proven to be effective, evaluated with 3500 sets of PD samples obtained experimentally. The overall pattern recognition accuracy of PD signals from five types of defect is 92.57%, an increase of 6.47% and 4.76% respectively compared with BPNN and SVM.

- The proposed CNN based PD pattern recognition is also effective for defect types with high similarity (For example, type 2 and type 3 PD). The PD pattern recognition accuracy of CNN for defect type 2 was 10.00% and 5.87% better than BPNN and SVM respectively. The accuracy of CNN for defect type 3 was 16.11% and 12.60% better than BPNN and SVM respectively.

- As outlined in the paper, the structure of the CNN – the number of network layers, the convolutional kernel size, the activation function and the pooling method applied – affects the accuracy of the CNN based PD pattern recognition method.

- As the number of network layers increases, the ability of CNN to correct the reference data increases. Based on the dataset used in this work, a 7-layer CNN has the highest overall PD pattern recognition accuracy and the highest accuracy for each type of defect.

- The size of the convolutional kernel has significant impact on the CNN based PD pattern recognition accuracy. For this dataset a convolutional kernel size of $1 \times 4$ is shown to be most accurate and to have moderate computational cost.

- Based on the dataset, the overall pattern recognition accuracy for the five types of PD signals based on CNN with swish was better by 3.43%, 2.56% and 0.95%, compared with CNNs with tanh, sigmoid, and relu, respectively.
The overall accuracy rate for PD pattern recognition based on CNN with RAP was better by 0.57% and 0.76%, compared with CNNs with max pooling and average pooling systems respectively. All studies in the paper were based on laboratory simulated artificial defects because for most on-line PD monitoring systems used in on-site PD testing, it is difficult to correlate PD signals with corresponding cable defect types without subsequent detailed forensic investigation of the defect topology after the insulation system has been removed from service. In cases where the defect develops into a fault, mechanical degradation will, in many situations, preclude such analysis. Whether the pattern recognition methods trained on artificial defects could be applied to actual defects will require further analysis of online monitoring data.

The proposed CNN based PD pattern recognition shows remarkable improvement for PD type 2 and type 3, but the pattern recognition accuracies are still not too high, 88.57% for PD type 2 and 78.10% for PD type 3. This was due to PD type 2 and type 3 being topologically similar. Further research, such as new PD feature construction or new pattern recognition methods, will continue to be researched to further improve pattern recognition accuracy.

REFERENCES


Donald M. Hepburn received his B.A. (Hons) from the Open University in 1987 and the Ph.D. degree from Glasgow Caledonian University (GCU) in 1994. He has many years of industrial research experience and has been involved in research into high voltage insulation systems at GCU for over 20 years. His research interests cover monitoring of chemical changes to insulation materials, application of electrical, acoustic and RF monitoring equipment to HV components and application of advanced digital signal processing to information from the monitoring techniques. He is a Senior Lecturer at GCU and is involved in industrial and academic research projects.

Alistair J. Reid (M’11–SM’15) graduated from the University of Strathclyde, UK, in 2004 with a B.Eng. (Hons) degree in Electrical and Mechanical Engineering and received the Ph.D. degree in 2007 for research on partial discharge monitoring. Subsequently, he has worked as a Post-Doctoral Researcher at the University of Strathclyde, a Research Fellow at Glasgow Caledonian University and a Visiting Research Associate at California Institute of Technology. He is presently a Lecturer in the School of Engineering at Cardiff University. His research interests include diagnostic monitoring techniques for high voltage systems, HVDC, gas mixtures, sensors, and energy harvesting. Alistair is a Senior Member of the IEEE.

Martin D. Judd (M’02–SM’04) is Technical Director of High Frequency Diagnostics based in Glasgow, Scotland. He was born in Salford, England in 1963 and graduated from the University of Hull in 1985 with a 1st Class (Hons) degree in Electronic Engineering, after which he gained 8 years of industrial experience with both Marconi Electronic Devices and EEV Ltd. Martin received his PhD degree from the University of Strathclyde in 1996 for research into the excitation of UHF signals by partial discharges in gas insulated switchgear. He has worked extensively on UHF partial discharge location techniques for power transformers and was latterly Professor of High Voltage Technologies at the University of Strathclyde, where he managed the High Voltage Research Laboratory. In 2014, he founded High Frequency Diagnostics, a specialist consultancy business that works in partnership with companies developing new electromagnetic wave sensor technologies and applications.

W. H. Siew (M’95) is a Reader in the Department of Electronic & Electrical Engineering, University of Strathclyde, Glasgow, Scotland. He is a triple alumnus of the University of Strathclyde with degrees of B.Sc. (Hons) in electronic & electrical engineering; Ph.D. in electronic & electrical engineering; and Master of Business Administration. His areas of research interest include large systems electromagnetic compatibility; cable diagnostics; lightning protection; and wireless sensing systems. He is Convener of the CIGRE WG C4.208 and a member of the Technical Advisory Panel for the IET Professional Network on Electromagnetics. He is also a member of IEEE TC7. He is a Chartered Engineer and an MIEE.