Yeo, J.; Jin, H.; Rodrigo Mor, A.; Yuen, C.; Tushar, W.; Saha, TK.; Seng Ng, C. (2021). Identification of Partial Discharge Through Cable-Specific Adaption and Neural Network Ensemble. IEEE Transactions on Power Delivery. 1-10. https://doi.org/10.1109/TPWRD.2021.3093670

Identification of Partial Discharge Through Cable-Specific Adaption and Neural Network Ensemble

J. Yeo, *Student Member, IEEE*, H. Jin, A. Rodrigo Mor, C. Yuen, *Fellow, IEEE,*
W. Tushar, *Senior Member, IEEE*, T. K. Saha, *Fellow, IEEE,* and C. S. Ng, *Member, IEEE*

***Abstract*-- This paper proposes to administer a multi-step artificial intelligence approach with an ensemble of adaptive neural networks (NNs) trained on 50000 samples to identify partial discharge (PD) diagnostic measurements for in-service medium voltage (MV) power cables. To evaluate the performance of the algorithm, a case study was performed on cables deliberately selected to contain both uncomplicated measurements and disruptive irregularities representative of conditions during field testing. Experimental test results prove that the proposed cable-specific adaptation improves PD identification accuracy, with further increment through the NN ensembles. The main contribution of the approach is in both the cable-specific adaption and the NN ensemble being applied to MV cable field measurements.**

***Index Terms*-- partial discharge, neural networks, medium voltage cables**

I. Introduction

PARTIAL discharge (PD) is a standardised measurement test conducted on MV cables during commissioning stages to detect initial installation defects and later during maintenance stages to determine insulation degradation. Typically, to stress the insulation defects and incept discharge, multiple increasing voltage levels are applied. The large number of recorded measurements are often imperfect due to numerous prominent but non-exhaustive considerations – background noise, pollution by neighbouring sources, inherent attenuation and dispersion from the pulse propagation characteristics [1], recording errors, multiple discharge sites and ambiguous interpretations. The total number of the recordings combined with the waveform complexity encountered potentially create complications that requires intricate attention by domain experts – this arduous process poses as a dilemma to the operator during on-site testing, and creates substantial workload during off-site evaluation [2].

Recent years have seen a resurgence in the application of neural networks on PDs, undoubtfully from the breakthroughs seen by machine learning (ML) and deep learning (DL). Basic NN experiments on varying hyperparameters were performed on PD frequency [3] and power spectrum [4], feature extraction were performed on phase-resolved partial discharge (PRPD) patterns [5] - [6]. Experimental results all agree that NNs are highly accurate in PD pulse shape recognition under laboratory conditions. Inclusion of PD defects from the field [7] were attempted, which led to a decrease in evaluation accuracy; it can be deduce here that the noise-influenced measurements from field conditions must be given due considerations. Recognising the importance of identifying noise signals in practical applications, study into PRPD with noise signals was conducted [8], NNs were considered here for recognition, among data mining and ML methods, but was 10% less accurate than the proposed PD recognition technique. The contents of the works were then extended to [9] and [10], where in [9] further comparison of different manual feature selection were made, and in [10] features were extracted via wavelet transformation. Conclusion was drawn here that the NNs required large data set in order to perform well, but is useful for identification of PD. However, there is a general skepticism in the application of NNs due to the difficulty in explaining the good results achieved in automated diagnosis of PD. In an attempt to dissect the NN black box conundrum, works were performed to display visualisation of neuron activation [11].

The need for detailed research pertaining to how different noise level affect PD identification accuracy was acknowledged [12], as majority of the research done previously for PD identification was conducted with self-fabricated insulation defects within laboratory conditions. For on-site investigations, the increasing amount of electrical interference on PD measurements are well described [13] and current research focus has still largely been directed to applying discrete wavelet transformation for PD analysis in MV cables measured with high frequency current transformer (HFCT)

J. Yeo is with the Engineering Product Development pillar, Singapore University of Technology and Design, Singapore (joel_yeo@mymail.sutd.edu.sg).
H. Jin is with T Yeo Engineers Pte. Ltd., Singapore (hjf@tyeoengineers.com).
A. Rodrigo Mor is with Universitat Politècnica de València, Instituto de Tecnología Eléctrica, Valencia, Spain (arrodmor@ite.upv.es).
C. Yuen is with the Engineering Product Development pillar, Singapore University of Technology and Design, Singapore (yuenchau@mymail.sutd.edu.sg).
W. Tushar is with the Faculty of Engineering, Architecture and Information Technology, University of Queensland, Queensland, Australia (w.tushar@uq.edu.au).
T. K. Saha is with the Faculty of Engineering, Architecture and Information Technology, University of Queensland, Queensland, Australia (saha@itee.uq.edu.au).
C. S. Ng is with T Yeo Engineers Pte. Ltd., Singapore (csng@tyeoengineers.com).

coupled to live conductors. As described above, previous experimental investigations have been conducted on time-resolved partial discharge (TRPD) and PRPD confined largely to laboratory conditions, with either artificially recreated defects or slight inclusion of field measurements. Existing literature do not address complications faced by TRPD measurements obtained from in-service cables taken offline and excited with an external power source – there has not been a detailed study conducted on TRPD measurements gathered solely from in-service cables. While previous exploratory studies are critical for foundational understanding, they have difficulty translating to practical applications because they are unable to consider the consequences arising from field measurements.

Analysis in this paper is done in the time-resolved domain for field cable measurements, which enable us to examine the shape of the PD pulse in the nanosecond scale of the insulation defect and determine if the received measurement is PD or noise. It is evident from the literature reviewed that PD waveforms recorded during field measurements must be assessed carefully as they are measured under noise influences. It is necessary but laborious to build an exhaustive databank when utilising NNs as they are generally acknowledged to have the ability to excel on complex correlation task when given large amounts of well-selected data sets, though admittedly such a luxury is not always available when dealing with practical applications.

This paper investigates DL techniques and their applications for automatic PD identification. To address the potential data limitations, NN adaptation techniques are presented. To improve generalisation capabilities, the combination of multiple NNs is considered. In order to exhibit the applicability of this algorithm, a performance evaluation is conducted through a case study to explore the success in overcoming these aforementioned impediments. The framework for the multi-step methodology is as follows, creation of a PD databank used to train several NNs, adaptation of the individual NNs through retraining on the measured calibration pulse of the measuring system of the cable under evaluation, formulation of a prediction and decision through an ensemble.

The structure of the paper is as follows – Section 2 introduces the background of the study, Section 3 proposes the algorithm of the experimental work, Section 4 describes the results of the methodology and finally Section 5 summarises for conclusion.

A. Partial Discharge Fundamentals

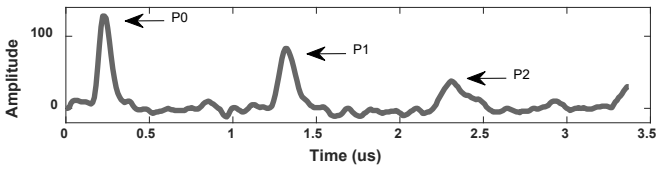

Fig. 1. Example of PD measurement in cable.

Fig. 1 is the classical signature of a PD in an offline power cable test that shows three decaying peaks – the triggering pulse (P0), the first reflection (P1) and the second reflection (P2); the time difference between the first reflection and the triggering pulse denotes the discharge location through the travelling wave theory. Normally, because of the signal attenuation and dispersion, the pulse waveforms are distorted in such a way that the amplitudes of the three consecutive pulses decrease and the pulses elongate. In some cables, depending upon the internal and external noise influences, mismatch impedances from joints, signal attenuation and dispersion and total cable length, it is possible to measure more complex waveforms that are a challenge for the expert to analyse.

It is understood and required by IEC 60270 that calibration of the PD measurement system is needed before any test and is done by a charge calibrator. Using a calibrator, a pulse of known charge is injected into the circuit. The injected charge value is reduced to the minimum detectable level which determines the sensitivity check of the measurement. By estimating the damping seen from the energy loss of the pulse and the successive pulse reflections, the required scale factor of the measurement is determined. This procedure is also used to determine the propagation time, and for the identification of the cable ends through the reflection due to impedance changes [14].

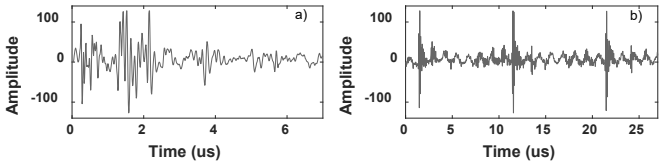

Fig. 2. Field noise measurement; (a) Random distributed noise, (b) Repetitive noise.

As mentioned previously, noise interference is an undeniable factor in field cable testing. It has nonstationary properties and could come in the simple form of small perturbations at zero-crossing, large messy spikes as shown on the top in Fig. 2a, or repetitive at random frequencies and magnitude as shown on the bottom in Fig. 2b. Given a noisy environment, it would not be unexpected that the recorded PD pulses are corrupted with noise. In order to recognise noise signature in the field, it is paramount that the noise sample data set is not filled with multiple copies of the same noise types but a wide variant of probable noisy conditions so that the NNs are able to generalise well and differentiate between PD and noise. Therefore, the recognition of PD pulses in noisy environments is a difficult task that needs an expert supervision to properly select a diverse and distinct data set in order for the NNs to react successfully to new data.

B. Neural Network Fundamentals

The basis for any NN is fundamentally similar; given an input it predicts an output. Information is fed forward to be multiplied with weight of the node and an output is predicted with the assistance of an activation function. The output for the sigmoid activation function can be defined as:

$$S(x) = \frac{e^x}{e^x+1} \quad (1)$$

where x is the input value and sigmoid function is used for binary interpretations.

Different groups of NNs are categorised by the movement

structure of the data. There are several elemental configurations that are understood to be applied for certain data structure. The convolutional neural network (CNN), recurrent neural network (RNN), and convolutional recurrent neural network (CRNN) are explored in this paper.

CNNs are the current state-of-the-art in computer vision and image recognition. The convolution property of the network captures spatial dependencies of the input data through the application of filters of adjustable sizes, with each hidden layer learning a corresponding feature of the data [15].

RNNs are the current state-of-the-art in processing of time-series data. The recurrent property of the network enables consideration of temporal influences during processing of sequential data through interconnection between nodes of each successive layer [16]. The vanishing gradient phenomena is formally addressed in Hocheriter's thesis [17] with the introduction of the Long Short-Term Memory (LSTM) cells. It is also possible to have Bidirectional Long Short-Term Memory (BILSTM) cells, which administers sequential information from both the past and future.

[18].

C. Data Acquisition

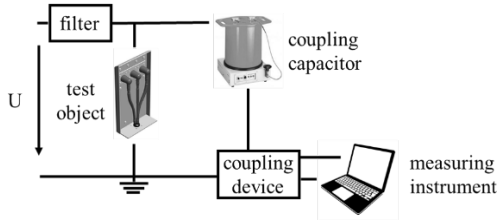

Fig. 3. On-site PD measurement setup.

In this study, PD testing and calibration of the measurement setup were in accordance with the industry standard IEC 60270 and is as shown in the Fig. 3.

TABLE I
Data Set Parameters

Parameters	
Sample size	47852
Type	XLPE
Length	50 m – 1500 m
Joints	2 – 17
Voltage	0.5, 1, 1.5, 1.7x U_0
Age	New and old mixed
PD vs. Noise	50%-50%

The parameters are as shown in the Table I. The data set consisted of 47852 samples equally distributed between PD and noise. The waveforms are carefully selected such that there are infrequent duplicates of the PD or noise signatures, this curation is to provide distinct diversity amongst the samples. Training set and validation set were divided with a split of 90% and 10% respectively from the data set. The samples were obtained from in-service cables taken offline and excited externally with a very low frequency test set. Tests were performed at different voltage levels, with the 1.7x U_0 as the maximum value; U_0 was 6.35 kV – this provides insight to the partial discharge inception and extinction voltage. Samples were only taken from XLPE cables to ensure homogeneity and tests were performed on different cable lengths consisting of different number of joints. The estimated localisation error is approximately 1% of the cable length and PD measurement range vary from 1 pC to 100 nC. The longest cable length in this data set was approximately 1.5 km. The sampling time was 10 ns, which provided a maximum bandwidth of 50 MHz and was sufficient for the cables measured.

Conventionally, the database is formed through a combination of all the PD and noise waveforms from the cables and divided randomly into three sets – which are the training set influencing the model, the validation set indirectly influencing the model, and the test set being purportedly providing an unbiased evaluation of the model fit on the data set. However, the disadvantage of such an arrangement is that the test set can contain traces of waveforms from the same cables being used to train model previously. While strictly speaking the model has not seen the data in the test set, it is highly probable that it has seen identical signatures during training. To counteract, the test set chosen in this paper is composed of five different cables that are not included in the database; this forms the case study. This subdivision also allows access into each individual background parameter of the cable for NN adaption purposes. These cables were deliberately selected because they were representative of both the typical measurements and disruptive irregularities found during field testing; the waveforms are exemplified in the next subsection.

III. Proposed Method

A. Algorithm Proposal

NNs are generally acknowledged to have the ability to excel on complex correlation task when given large amounts of well-selected data sets. The obvious solution would be to prepare a comprehensive databank comprising of different PD pulses and noise patterns from an assortment of different lengths. But the common constraint, admittedly, is that such a luxury is not always available due to resource limitations when dealing with data scarcity and environmental variability. Here exploration is done on how the NNs may overcome the aforementioned constraints through an initial meticulously formed databank, subsequently acclimatised to cable ground conditions by re-training with the measured calibration pulse of the measuring system, and finally supported by an ensemble.

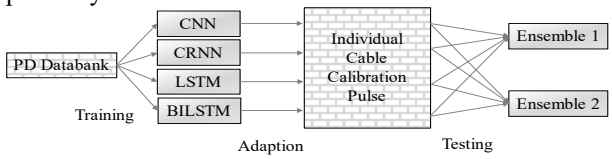

Fig. 4. Adaptive NN ensemble system.

Our adaptive NN ensemble system architecture is as shown in Fig. 4. The NN draws the data from the PD library and performs training till a desired state is reached, it is then adapted to the individual calibration pulse of the cable undergoing evaluation. Prediction of the test waveforms are made, and the results concatenated to the NN ensemble for the combined prediction. The main contribution of the approach is in both the cable-specific adaption and the NN ensemble being applied to MV cable

field measurements.

B. Cable-Specific Adaptation

In practice, as mentioned earlier, it is not unusual for the measurement to be distorted by noise originating both externally and internally [19]. External noise could arise from telecommunication transmission, arcing, power electronic switching pulses, and PD from the power system. Internal noise is a result of thermally induced current fluctuations from the measuring circuit. Instead of actively performing manual feature selection to separate the PD signal and suppress the noise, NNs in this proposed algorithm have been tasked to automatically extract features from the data set to identify PD that have been polluted by noise – examples of the highly challenging waveforms are as shown in Fig. 5. The authors have successfully proven the capability of this approach in previous works [2] [18] [13].

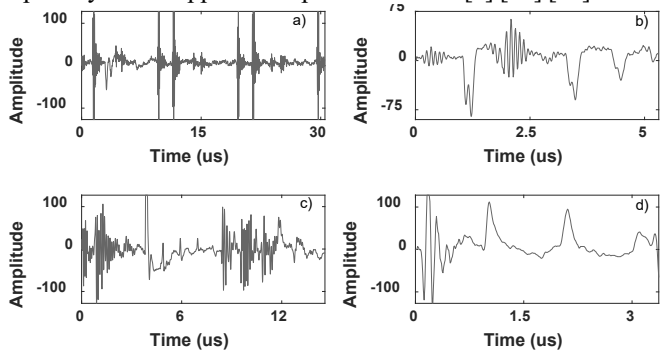

Fig. 5. (a) Small PD pulse with large repetitive noise, (b) Multiple PD pulses fused together, (c) Singular PD pulse with large noise, (d) Wrong triggered recording.

However, with consideration of all the aforementioned field complications, this paper proposes a more strategic approach for a comprehensive generalisation through the adaption of the NNs with readily accessible cable-specific context. The cable-specific context can be incorporated in the algorithm using the measured calibration pulse of the measuring system obtained during the cable calibration and the sensitivity check.

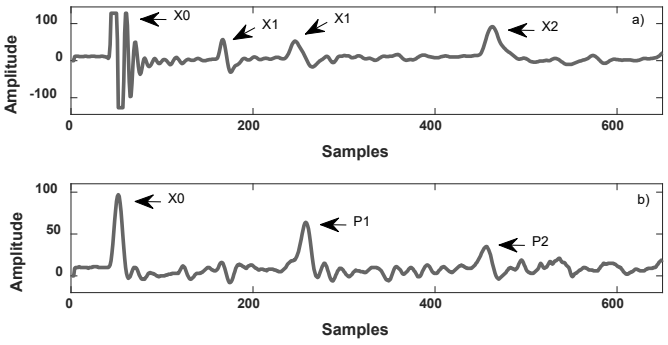

Fig. 6. (a) Calibration pulse with prominent over and undershoot, indicative of impedance change, (b) Ensuing PD measurements.

Apart from checking the sensitivity of the measurement system, the measured calibration pulse of the measuring system determined the cable length numerically through the amount of samples between pulses X0 and X2 from Fig. 6a when processed by the NNs. This waveform also reproduced probable ground conditions that occurs during the PD measurement, e.g. location of the triggering pulse with respect to the total recording, joint location, second reflection and ambient noise. Joint location can be observed with prominent over and undershoot as marked with pulse X1 in Fig. 6a, which is typical of impedance change – this information aids the machine in recognition of prospective discharge spots by providing computational clues as to where the discharge location might be; cable joints are recognised to be common insulation weak spots that causes PD. Comparing Fig. 6a and Fig. 6b, the calibration triggering pulse X0 is in the same position as the PD triggering pulse P0. The calibration pulse reflecting at the end of the cable is the pulse X2, which is in the same position than the PD second reflection pulse P2. Regarding the position of the joints, the discharge location of pulse P1 is in the same position than the second joint location indicated by the pulse X1 in Fig. 6a. This visibly illustrates the correlation between the calibration pulse and PD measurement. These factors reinforce the advantages derived from utilising cable-specific context.

The adaption was performed by re-training only the classifier layer of the trained NNs with the measured calibration pulse of the measuring system – this adaption is unique to each cable evaluation. The proposed technique was motivated by the principles of transfer learning and previous works done on speaker adaption on speech recognition [20]. The results of the effectiveness for the adaption will be presented on the case study cables through comparison of the statistical analysis both before and after adaption; it is described through the performance of the individual NNs and the ensembles.

C. Neural Network Ensemble

The basic idea of an ensemble is to classify the given input through consensus classification from multiple NNs. This can be achieved in two different ways through equation 1 – the output of the sigmoid activation function. The sigmoid activation function yields a decimated value between 0 and 1; during binary classification, a threshold value is set for conversion of the output decimated value. Two different ensembles are proposed, where each ensemble will be made up of two different NNs giving a total of six different models for each ensemble.

Ensemble 1 analyses the binary result for the prediction of two different models – PD or noise. When there is a difference in prediction between models, the sample is removed and deferred to be examined by the domain expert. Ideally, only a minimal amount of complicated waveforms should be conveyed for external analysis. This proposed method expects the NNs to be highly accurate, as otherwise a large number of waveforms would be deferred and is not the intention. The authors have successfully proven the capability of this approach in a previous work [18].

Ensemble 2 sums the output decimated value from the activation function of two different models and sets an accordingly higher threshold value for binary conversion. As a NN may perform better in analysing certain waveforms signatures, it can assist another NN with uncertain prediction and compensate for the shortcomings. For instance, in analysing the same PD waveform, the threshold for PD recognition is set at 0.5, given the boundaries of 0 to 1 from the sigmoid function in (1). The first NN being hesitant could have tabulated a result of 0.4 whereas the second NN could have more confidently computed a result of 0.7. Individually, this would have resulted in a noise and PD classification respectively for the two NNs. In the proposal of *Ensemble 2*, the threshold herein would be increased from 0.5 to 1 due to the summated decimal output. With

the proposed summation, the second NN is able to edge the combined evaluation across the threshold and correctly label the waveform as PD.

The results of the effectiveness for the ensembles will be presented on the case study cables through comparison of the statistical analysis between the individual NNs and the ensembles. Pertinent to analysis of *Ensemble 1* would be the amount of waveforms excluded during evaluation, which should be a minimal number of complicated waveforms. For *Ensemble 2*, observation would be made to ascertain the likelihood of a more confident NN providing compensation for complicated waveforms.

IV. Experimental Results

A. Case Study

The five cables under the case study were not included in the training data set. They were deliberately selected because they contained both uncomplicated measurements and disruptive irregularities representative of field conditions which are difficult to organically replicate in the laboratory – plotted in Fig. 7 to Fig. *11* are either noise signatures, PD pulses or a combination of both.

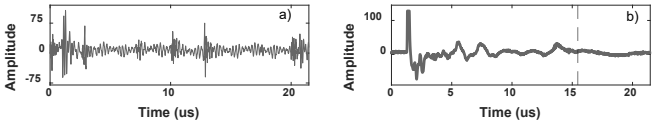

Fig. 7. *Cable 1*; (a) Noise waveform (left), (b) Single peak PD (right).

Cable 1 had 775 recordings and was slightly over 1000 m with more than 20 joints; the end of the cable is denoted with the dotted vertical line on Fig. 7b. It contained pulsating noise as shown in Fig. 7a, this large noise influence is present in the PD recordings as shown in the mild oscillations of Fig. 7b. This is a common and unmistakable singular PD peak waveform, typical in longer cables with multiple joints.

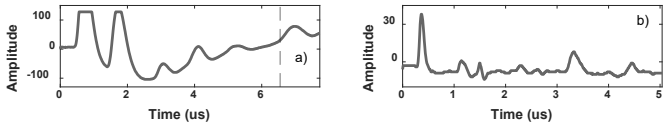

Fig. 8. *Cable 2*; (a) Two peak PD (left), (b) Interrupted PD recording (right).

Cable 2 had 507 recordings and was 480 m long. As shown on Fig. 8a, there are two large peaks, with the latter being the discharge site location; the end of the cable is denoted with the dotted vertical line. As shown in Fig. 8b, while it is plausible that the second largest peak is the first reflection, it is difficult to be certain as this is an incomplete recording – interruption was at approximately the 5 us mark. In both plots, it is also possible to see evidence of multiple discharge sites making it is hard to distinguish between noise influence and multiple discharge sites.

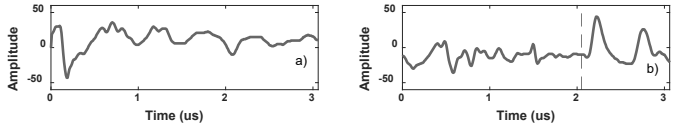

Fig. 9. *Cable 3*; (a) Noise waveform (left), (b) Wrongly triggered PD recording (right).

Cable 3 had 546 recordings and was 150 m long. As shown on Fig. 9a, it contained noise in both high and low frequencies, and a DC bias; this same noise signature can be seen at the start of Fig. 9b. Also shown in Fig. 9b, wrong trigger of the PD recording caused the two peaks to be at the end of the waveform instead of the start, the end of the cable is denoted with the dotted vertical line.

Cable 4 had 1828 recordings, was 150 m long, and contained discharges in two locations, as shown with the different PD plots in Fig. 10a and Fig. 10b. In Fig. 10a, wrongly triggered recording of a singular pulse PD recording can be seen, the noise perturbation coupled with DC offset caused it to have an amplitude similar to the PD pulse. In Fig. 10b the end of the cable is denoted with the dotted vertical line, it can be seen that there is excessive recording. Combination of the erroneous measurement poses as an obstacle because this causes recordings that are uncommon and therefore biased in training class – samples in the PD databank of such peculiarity are not statistically represented.

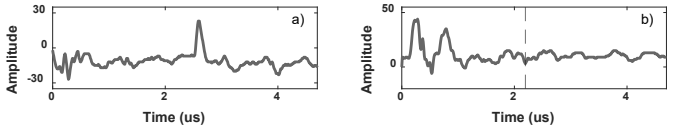

Fig. 10. *Cable 4*; (a) Early trigger singular PD pulse (left), (b) Excessive recording two peak PD (right).

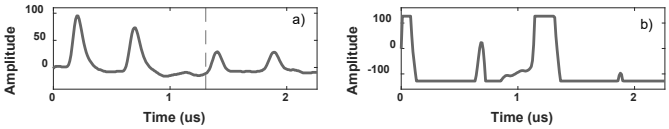

Fig. 11. *Cable 5*; (a) Excessive recording four peak PD pulse (left), (b) Noise waveform (right).

Cable 5 had 252 recordings, was 100 m long, and similarly suffered from excessive recording, but with a different complication. The end of the cable is denoted with the dotted vertical line as shown in Fig. 11a, the unnecessary recording resulted in four PD peaks – an understandably abnormal activity. As shown in Fig. 11b, at first glance the waveform looks to contain peaks identical to clipping of PD along with multiple discharge sites comparable to Fig. 8, but it is actually misleading oscillations if the amplitude is observed carefully. This noise signature can be extremely confusing to the NNs as these waveforms resemble clipped PD.

B. Test Parameters

The hyperparameters for the various models were chosen after extensive comparison. For the CNN – size of filters, convolution and pooling layers were examined; simpler architecture generated better performance. The CNN model contained two convolutional layers of four filters, average pooling layer of four units was placed after each convolutional layer. It is then connected to a dense layer and finally the output layer. For the CRNN – the model contained two convolutional layers of four filters, average pooling layer of four units was placed after each convolutional layer. The dense layer here is replaced with BILSTM cells and then connected to the output layer. For the LSTM and BILSTM – hidden layers and nodes were examined, slightly complicated architecture generated better performance. Both models had 4 hidden layers and 512 nodes in each layer. Dropout of 0.2 was placed on the visible layer of all four NNs.

The data set was divided into training and validation set with a split of 90% and 10% respectively; the case study had a total of 3908 recordings from 5 cables. Adam optimisation algorithm [21] was used for the gradient descent algorithm and the loss function was calculated with the mean squared error. The training was done on a single GPU setup with Keras [22] and Tensorflow 2 [23] as the backend.

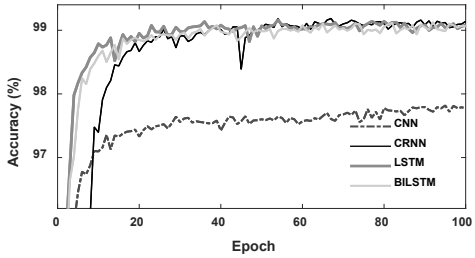

Fig. 12. Validation set accuracy.

It is understood that DL training is a stochastic process that could yield a different result each iteration. For an unbiased impression of the model general performance, it is necessary to observe the average results over an arbitrary amount of iterations. The training of the model was performed over 10 iterations, with 100 epoch each iteration. Models could be saved on either highest validation accuracy or lowest validation loss; the highest validation accuracy model was chosen as it gave better performing weights when compared. Validation accuracy averaged over 10 iterations are as shown in Fig. 12, the convergence of the NNs can be seen. Majority of the models fluctuate close to the 99% mark, with the CNN slightly below 98%.

After the model is obtained, for evaluation of each specific cable, the top layer before the classifier is re-trained with the measured calibration pulse of the measuring system through stochastic gradient descent for 10 epochs. The adaption training process is rapid and accepts either polarity for the calibration pulse.

TABLE II
Individual neural network success ration of PD identification.

		CNN	CRNN	LSTM	BILSTM
Cable 1	Acc (%)	98.97 (+0.65)	99.87(-)	100 (+0.26)	100 (+0.26)
	Pre (%)	99.63 (-0.67)	99.67 (-)	100 (-)	100 (-)
	Rec (%)	98.01 (+2.31)	100 (-)	100 (+0.66)	100 (+0.66)
Cable 2	Acc (%)	93.49 (+1.58)	82.84 (+1.58)	86.79 (+5.72)	86 (+2.96)
	Pre (%)	97.08 (-0.6)	93.35 (-4.51)	94.05 (-5.12)	92.81 (-4.86)
	Rec (%)	95.58 (+2.44)	86.95 (+6.2)	90.93 (+11.51)	91.37 (+9.29)
Cable 3	Acc (%)	92.49 (+1.83)	91.76 (+0.18)	94.14 (+1.47)	94.69 (+1.47)
	Pre (%)	94.7 (-0.8)	95.37 (-0.74)	94.96 (-0.64)	96.24 (-0.12)
	Rec (%)	97.18 (+3.03)	95.56 (+1)	98.79 (+2.42)	97.98 (+1.81)
Cable 4	Acc (%)	92.29 (+9.85)	96.44 (+8.26)	97.59 (+5.85)	97.98 (+10.78)
	Pre (%)	98.63 (-0.96)	99.71 (-0.16)	99.77 (-0.17)	99.1 (-0.71)
	Rec (%)	93.34 (+11.11)	96.62 (+8.69)	97.74 (+6.2)	98.82 (+11.85)
Cable 5	Acc (%)	98.41 (+3.57)	95.24 (+14.29)	89.29 (+4.37)	90.48 (+7.94)
	Pre (%)	98.7 (-0.39)	95.4 (-1.01)	93.53 (-2.61)	94.76 (+0.07)
	Rec (%)	99.56 (+4.36)	99.56 (+17.46)	94.76 (+7.86)	94.76 (+9.17)
Average	Acc (%)	94.19 (+5.42)	94.63 (+5.02)	95.65 (+4.02)	95.88 (+6.19)
	Pre (%)	97.85 (-0.77)	97.86 (-0.87)	97.8 (-1.07)	97.57 (-1.17)
	Rec (%)	95.11 (+7.38)	95.63 (+6.98)	96.96 (+5.97)	97.48 (+8.73)

The performance of the individual NNs and the ensemble on evaluating the cables will be given in the succeeding tables. The results of the adapted NNs are displayed, with the difference between non-adapted NNs given beside in a bracket; plus indicates improvement whereas minus implies deterioration. Apart from the accuracy (Acc), two other standard statistical evaluation parameters are displayed for further measures which are precision (Pre) and recall (Rec). Precision is the ability to identify only relevant data points and recall is the ability to find all relevant cases in a data set. For *Ensemble 1*, an additional parameter (Exc) depicting the amount of excluded samples is included in this table. The best performing model for the cables individually and averaged is highlighted. Discussion will be on the general overview of the results, emphasising on the best performing model that was enhanced by the adaptation and improved by the ensemble. Good performance in *Cable 1* is expected as this was an uncomplicated measurement. The enhanced generalisation capability of the proposed algorithm on *Cables 2, 3* and *4* is of further interest.

C. *Individual Neural Network Results*

Shown in Table II is the performance of the individual NNs.

Cable 1: All the NNs were capable of identifying the case of noise and singular peak PD in the cable. After adaption, slight improvement was observed.

Cable 2: The accuracy of all the NNs except the CNN was below 90% and sees overall improvement after adaptation. This file consisted of multiple discharge site, and distortion of the PD pulse due to reflections.

Cable 3: The accuracy for all NNs was greater than 90% and sees marginal improvement after adaptation. Due to the dispersion in the frequency domain, the width of the PD pulse has increased slightly causing it to resemble noise signatures of the cable with both high and low frequencies. Despite the resemblance, the NNs generally were able to differentiate and identify the PD.

Cable 4: Excessive recording by the system results in misleading interpretations. After adaptation the confusion was corrected, all NNs saw significant increase in accuracy and decrease in the false negatives.

TABLE III
Confusion Matrix of CRNN for *Cable 5*

Before adaptation		Prediction	
		Noise	PD
Label	Noise	16	7
	PD	41	188

After adaptation		Prediction	
		Noise	PD
Label	Noise	12	11
	PD	1	228

Cable 5: Four PD pulses appear in these samples due to excessive recording. To the best of the author's recollection this PD signature is not commonly found in the data set, which causes imbalanced data and the inability to identify the waveforms. Shown in Table III is the confusion matrix for the CRNN model, the results before adaptation are on the left. After adaption, significant improvement can be observed as shown, with the false negative again greatly reduced and the true positive increased accordingly.

On the average after adaptation, the BILSTM was the best performing model, followed by the LSTM, CRNN and then CNN.

TABLE IV
Ensemble neural network success ration of PD identification.

			CNN BILSTM	CRNN BILSTM	LSTM BILSTM	CNN LSTM	CRNN LSTM	CRNN CNN
Cable 1	Ens. 1	Acc (%)	100 (-)	100 (-)	100 (+0.26)	100 (-)	100 (-)	100 (-)
		Pre (%)	100 (-)	100 (-)	100 (-)	100 (-)	100 (-)	100 (-)
		Rec (%)	100 (-)	100 (-)	100 (+0.66)	100 (-)	100 (-)	100 (-)
		Exc	8 (+7)	1 (+2)	0 (-)	8 (+7)	1 (2)	9 (+5)
	Ens. 2	Acc (%)	99.74 (+0.51)	99.87 (-0.13)	100 (+0.26)	99.74 (+0.51)	99.87 (-)	99.61 (+0.38)
		Pre (%)	99.34 (-0.66)	99.67 (-0.33)	100 (-)	99.34 (-0.66)	99.67 (-)	99.02 (-0.98)
		Rec (%)	100 (+1.99)	100 (-)	100 (-0.66)	100 (+1.99)	100 (-)	100 (+1.99)
Cable 2	Ens. 1	Acc (%)	95.9 (+2.32)	87.53 (+3.85)	87.89 (+3.97)	95.94 (+3.12)	88.29 (+5.24)	94.69 (+1.65)
		Pre (%)	97.32 (-1.86)	94.9 (-4)	94.9 (-4.25)	97.56 (-1.88)	96.06 (-3.37)	97.47 (-1.71)
		Rec (%)	98.28 (+4.69)	91.36 (+8.64)	91.7 (+9.07)	98.03 (+5.54)	91.12 (+9.49)	96.74 (+3.7)
		Exc	68 (+3)	42 (-19)	20 (+8)	64 (+11)	46 (-17)	74 (+2)
	Ens. 2	Acc (%)	94.48 (+3.75)	85.01 (+3.55)	86.59 (+5.13)	93.89 (+4.34)	85.6 (+4.53)	92.9 (+2.96)
		Pre (%)	95.69 (-2.64)	92.15 (-6.23)	93.44 (-5.2)	95.66 (-2.88)	92.78 (-5.85)	95.22 (-3.09)
		Rec (%)	98.23 (+7.08)	90.93 (+10.4)	91.37 (+11.06)	97.57 (+7.97)	90.93 (+11.06)	96.9 (+6.63)
Cable 3	Ens. 1	Acc (%)	96.67 (+0.69)	95.38 (+0.6)	95.67 (+1.34)	95.92 (+0.87)	95.01 (+1.28)	93.89 (+0.3)
		Pre (%)	96.73 (-0.06)	96.11 (-0.33)	96.22 (-0.1)	96.38 (-0.24)	95.75 (-0.32)	95.93 (-0.69)
		Rec (%)	99.79 (+0.88)	98.95 (+1.08)	99.18 (+1.66)	99.38 (+1.31)	98.95 (+1.87)	97.52 (+1.11)
		Exc	36 (+12)	26 (+3)	15 (+2)	31 (+10)	25 (-5)	22 (+9)
	Ens. 2	Acc (%)	94.32 (+1.28)	94.69 (+3.11)	94.14 (+1.1)	93.22 (+0.55)	93.96 (+2.02)	92.12 (+0.73)
		Pre (%)	94.29 (-1.88)	95.52 (-0.4)	94.96 (-1.02)	93.55 (-2.6)	94.95 (-0.98)	94.15 (-1.57)
		Rec (%)	99.8 (+3.63)	98.79 (+4.03)	98.79 (+2.42)	99.4 (+3.63)	98.59 (+3.43)	97.38 (+2.62)
Cable 4	Ens. 1	Acc (%)	99.28 (+1.31)	98.7 (+9.01)	98.77 (+6.5)	99.4 (+1.65)	98.1 (+5.51)	99.21 (+2.2)
		Pre (%)	99.57 (-0.27)	99.71 (-0.16)	99.77 (-0.16)	99.75 (-0.18)	99.77 (-0.16)	99.75 (-0.09)
		Rec (%)	99.7 (+1.66)	98.96 (+9.48)	98.97 (+6.89)	99.63 (+1.89)	98.27 (+5.86)	99.44 (+2.39)
		Exc	154 (+347)	56 (+36)	37 (+84)	165 (+243)	41 (+72)	180 (+275)
	Ens. 2	Acc (%)	98.36 (+5.96)	98.03 (+9.63)	98.09 (+7.06)	98.41 (+8.37)	97.81 (+7.16)	98.25 (+6.57)
		Pre (%)	98.66 (-1.16)	99.71 (-0.1)	99.71 (-0.1)	98.93 (-0.94)	99.71 (-0.17)	98.88 (-1)
		Rec (%)	99.66 (+7.33)	98.25 (+10.04)	98.31 (+7.39)	99.44 (+9.59)	98.03 (+7.56)	99.32 (+7.78)
Cable 5	Ens. 1	Acc (%)	99.12 (+4.19)	96.15 (+8.41)	91.36 (+4.72)	99.11 (+4.44)	95.71 (+8.98)	98.76 (+3.5)
		Pre (%)	99.08 (+0.1)	96.44 (+0.31)	94.69 (-1.25)	99.08 (+0.08)	96.02 (-0.26)	98.70 (-0.77)
		Rec (%)	100 (+4.43)	99.54 (+9.38)	95.96 (+6.81)	100 (+4.81)	99.54 (+10.81)	100 (+4.62)
		Exc	24 (+11)	18 (+22)	9 (+11)	27 (-)	19 (+7)	10 (+31)
	Ens. 2	Acc (%)	95.24 (+2.78)	94.05 (+7.94)	90.87 (+7.93)	93.65 (+2.38)	93.25 (+7.93)	96.34 (+5.95)
		Pre (%)	95.4 (-3.67)	94.58 (-1.61)	94.02 (-0.69)	94.19 (-4.86)	93.8 (-2.35)	96.22 (-2.82)
		Rec (%)	99.56 (+6.98)	99.13 (+10.92)	96.07 (+10.04)	99.13 (+7.86)	99.13 (+11.79)	100 (+9.61)
Average	Ens. 1	Acc (%)	98.65 (+1.3)	96.97 (+5.33)	96.73 (+4.04)	98.59 (+1.58)	96.72 (+4.01)	98.03 (+1.5)
		Pre (%)	98.82 (-0.32)	98.31 (-0.63)	98.22 (-0.74)	98.88 (-0.33)	98.41 (-0.58)	98.78 (-0.38)
		Rec (%)	99.57 (+2.02)	98.06 (+7.19)	97.88 (+5.74)	99.44 (+2.35)	97.66 (+5.53)	98.87 (+2.32)
		Exc	290 (+380)	143 (+44)	81 (+105)	295 (+271)	132 (+59)	295 (+322)
	Ens. 2	Acc (%)	97.36 (+3.74)	95.98 (+5.89)	95.96 (+4.68)	97.03 (+4.79)	95.8 (+4.73)	96.85 (+4.02)
		Pre (%)	97.38 (-1.61)	97.63 (-1.11)	97.72 (-0.98)	97.29 (-1.75)	97.57 (-1.22)	97.45 (-1.49)
		Rec (%)	99.51 (+6.21)	99.45 (+6.15)	97.42 (+6.7)	99.2 (+7.63)	97.39 (+7.01)	98.8 (+6.43)

Before adaption, the BILSTM fared slightly poorer than the LSTM. All models improved after the adaption and were greater than 94%, slight decrement in precision scores were constantly seen. Adaption of the NNs proved to be useful for increasing identification accuracy.

D. Ensemble Results

Shown in Table IV is the performance of *Ensemble 1* and *2*. The results are as follows:

Cable 1: All combinations in *Ensemble 1* except for LSTM-BILSTM performed better than *Ensemble 2*, deferring a small percentage of uncertain waveforms. The general overall accuracy of both ensemble is consistently high. For *Ensemble 1*, results for most waveforms had no identification mistakes before adaption and were at 100% accuracy. The LSTM-BILSTM combination achieved the same results after a slight increment from adaptation. Overall, the amount of waveforms deferred for external analysis were less than 1% of the total evaluated, with lesser files being deferred after adaptation. For *Ensemble 2*, adaptation of the ensemble generally gave slight improvements, except for the CRNN-BILSTM which saw a decrement and CRNN-LSTM which had no improvement; all the combinations were close to 100% accuracy. Except for the LSTM-BILSTM, all the combinations performed slightly lower than *Ensemble 1*. As this was a relatively uncomplicated measurement, the best performing ensemble and individual NN performed equally as good.

Cable 2: CNN-LSTM in *Ensemble 1* marginally outperformed CNN-BILSTM in *Ensemble 1* but defers almost 13% of the waveforms. The general overall accuracy has a fluctuation in the scores, demonstrating the capabilities of some ensembles over others. For *Ensemble 1* and *2*, adaptation increased the overall ensemble accuracy. The two best combinations were the CNN paired with either the BILSTM or the LSTM, with both yielding comparable performance. Amongst *Ensembles 1*, the amount of waveforms deferred for external analysis ranged from 4% to 14%. Comparing the ensembles with individual NNs, the best results from both the adapted ensemble were better than individual NN – the CNN.

Cable 3: Results of the statistical analysis in *Ensemble 1* and *2* were consistent throughout the various models and had a low standard deviation; the two ensembles were comparable in performance. For *Ensemble 1*, adaptation provided a marginal increment, whereas slightly more contribution was seen in *Ensemble 2*. Almost 2% of increment in accuracy was achieved in *Ensemble 1*, in comparison to *Ensemble 2*. The best performing model of *Ensemble 2*, the CRNN-BILSTM, had the same accuracy as the best performing individual NN – the BILSTM. The amount of files deferred for external analysis in *Ensemble 1* were less than 10%, except for the CRNN-LSTM combination, lesser files were deferred after adaptation.

Cable 4: Results of the statistical analysis in *Ensemble 1* and *2* were close to 100% and consistent throughout the various models with a low standard deviation; the two ensembles were comparable in performance. For *Ensemble 1*, adaptation provided significant improvement in four of the models. The amount of files deferred for external analysis were less than 1% of the total evaluated, with lesser files being deferred after adaptation. For *Ensemble 2*, akin to *Ensemble 1*, adaptation of the ensembles saw large increase in accuracies throughout, with all models performing approximately at 98%. Comparing the ensembles with individual NNs, the best results from both the adapted ensemble were better than individual NN – the BILSTM.

Cable 5: Large variance is observed in the results from *Ensemble 1* and *2*. The PD signature is not commonly found in the data set and as a result caused an imbalance in the training data set for recognition. For *Ensemble 1*, adapted ensemble results from the CNN combinations performed closed to 100% accuracy, good amount of overall increment was seen by the combinations after the adaptation. The amount of files deferred for external analysis was less than 10% of the total evaluation, with lesser files deferred after adaptation. For *Ensemble 2*, adaptation of the ensembles helped the models to improve and achieved between 90% to 96%. Comparing the ensembles with the individual NNs, CNN-BILSTM from *Ensemble 1* is the best performing model, followed by the individual CNN outperforming the best performing model of *Ensemble 2* – the CRNN-CNN. This is the only cable where the *Ensemble 2* did not outperform an individual NN.

On the average for *Ensemble 1*, the results agree with the performance from the individual models, where the CNN and BILSTM models were better. All the models improved after adaption and the accuracies were greater than 95%. Large majority of the files had a reduction in precision scores after adaptation. The amount of deferred samples after adaptation was reduced tremendously. On the average for *Ensemble 2*, the CNN and BILSTM was similarly the superior model. All the ensembles improved after adaptation and accuracies ranged from 94% to 96%. Comparing the performance in evaluating the cables between the ensembles, *Ensemble 1* outperforms *Ensemble 2* for all models; this is expected as waveforms were excluded for external analysis. However, the best performing model for *Ensemble 2* is consistently comparable to *Ensemble 1*, with only a marginal 1% to 3% difference between them. The best performing individual NN model, the BILSTM, did not outperform either of the ensembles. This clearly validates that ensemble NNs are useful in increasing identification accuracy. Very much like the individual NNs, adaption of the NNs causes a decrement in precision scores, but proved also to be useful for increasing identification accuracy.

E. Further Analysis

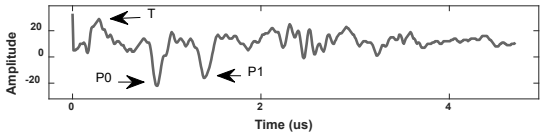

Fig. 13. *Cable 4*; Example of complicated waveform.

Shown in Fig. 13 is an example of a complicated PD measurement from *Cable 4*. While the general PD characteristics can be established with the triggering pulse and first reflection denoted by P0 and P1 respectively, the signal to noise ratio is not significantly large. The measurement was triggered incorrectly by a large spike denoted T at the start of the recording, had a DC offset and contained high and low frequency noise. This results in T being larger in amplitude than P0 and resembles misleading noise oscillations observed in Fig. 9a. The CNN-LSTM ensemble model will be discussed as it produced the highest accuracies in both ensembles. Before adaptation, both the CNN and CRNN labeled the waveform as noise, whilst the LSTM and BILSTM labeled the waveform as PD. From (1), the CNN predicted a value of 0.301, whereas the LSTM predicted 0.99. Through the proposed methodology of *Ensemble 2*, the more confident NN can lead the evaluation in the right direction – this reinforces the utilisation of the NN ensembles. After adaption, both the CNN and CRNN were able to correctly label the waveform as PD.

TABLE V
Confusion Matrix of CNN for *Cable 4*.

Before Adaptation		Prediction	
		Noise	PD
Label	Noise	49	6
	PD	315	1458

After Adaption		Prediction	
		Noise	PD
Label	Noise	32	23
	PD	118	1655

The confusion matrix for the CNN is as given in Table V, shown on the left is the results before adaptation. After adaptation, the true positives have increased significantly whilst the false negative decreased in accordance – this unmistakably reinforces the advantage of adaptation. The two demonstrated factors from the adaption and ensemble positively exemplify the advantages of the proposed algorithm.

V. Conclusion

This paper investigates identification of PD measurements from in-service cables and proposes an algorithm that is applicable to TRPD cable measurements. The presented methodology is able to identify highly challenging PD waveforms that have been polluted by noise, this was demonstrated on 5 case study cables which were independent from the training data set.

Individual NN validation set results were remarkably high, which represents the ability of the NNs in identifying the waveforms. However, the five case study files before adaption do not completely reflect the same degree of accuracy seen previously. Such varying results are not uncommon representation of the generalisation competencies for field measurements. While good performance for straightforward measurements such as *Cable 1* are expected, the focus of the algorithm is on the complicated recordings as seen on *cables 2* to *5*.

Without the adaptation process, the average accuracy for the individual NNs were slightly below 90%. With the adaptation process, the individual NNs were able to increase the accuracy hoovering at 95%. This proves that the cable-specific adaptation has the ability to negate the uncertainty caused by unpredictable waveforms that are not included in the databank and is able to enhance the performance of highly trained NNs.

Ensemble NNs provides better overall performance compared to individual NNs; adaptation on ensemble NNs further improves the performance. Through *Ensemble 1*, an extremely high accuracy of 98.65% can be attained with the best performing combination CNN-BILSTM – with only 7.4% of the waveforms deferred to the operator. In *Ensemble 2* the best performing combination was similarly CNN-BILSTM, the accuracy was lower than *Ensemble 1*, but still remarkably successful at 97.36%. For both the ensembles the best performing combination was the CNN paired with the BILSTM, despite the CNN being the lowest performing individual model. This further ascertains the belief that NNs of differing principles, when used collectively, can negate shortcomings stemming from each individual NN type; the overall generalisation errors and variance of the prediction is also decreased significantly.

Drawback of adaptation is evident in the slight decrement of the overall precision; however, this is inconsequential considering the overall success. The successful results from adapting highly trained NNs to cable-specific conditions and evaluating thereafter with NN ensembles successfully demonstrates the capability of the algorithm. The proposed methodology distinctly displays a systematic approach to the challenges faced in field PD identification.

For future works, with respect to the increase in false positives, it is proposed to further analyse the predicted PD through consideration of the cumulative energy signal – as a PD signal contains energy, this feature extraction may prove to be a simple and efficient tool.

VI. References

[1] A. Rodrigo Mor, P. H. F. Morshuis and J. J. Smit, "Comparison of Charge Estimation Methods in Partial Discharge Cable Measurements," *IEEE Trans. Dielectr. Electr. Insul.,* vol. 22, no. 2, pp. 657-664, 2015.

[2] J. Yeo, H. Jin, C. Yuen and C. Ng, "Application of Recurrent Neural Network with Long Short-Term Memory Cells on Partial Discharge Identification," in *Jicable 10th Int. Conf. Insul. Pow. Cab.*, Versailles, 2019.

[3] A. Hirata, S. Nakata and Z. I. Kawasaki, "Toward Automatic Classification of Partial Discharge Sources With Neural Networks," *IEEE Trans. on Pow. Del.,* vol. 21, no. 1, pp. 526-527, 2006.

[4] T. Boczar, S. Borucki, A. Cichon and D. Zmarzly, "Application Possibilities of Artificial Neural Networks for Recognizing Partial Discharges Measured by the Acoustic Emission Method," *IEEE Trans. on Dielectr. Insul.,* vol. 16, no. 1, pp. 214-223, 2009.

[5] H. Ma, J. C. Chan, T. K. Saha and C. Ekanayake, "Pattern Recognition Techniques and Their Applications for Automatic Classification of Artificial Partial Discharge Sources," *IEEE. Trans. Dielectr. Electr. Insul.,* vol. 20, no. 2, pp. 468-478, 2013.

[6] M. Majidi, M. S. Fadali, M. Etezadi-Amoli and M. Oskuoee, "Partial Discharge Pattern Recognition via Sparse Representation and ANN," *IEEE Trans. on Dielectr. Insul.,* vol. 22, no. 2, pp. 1061-1070, 2015.

[7] H. Song, J. Dai, G. Sheng and X. Jiang, "GIS Partial Discharge Pattern Recognition via Deep Convolutional Neural Network Under Complex Data Source," *IEEE Trans. Dielectr. Electr. Insul.,* vol. 25, no. 2, pp. 678-685, 2018.

[8] X. Peng, J. Wen, Z. Li, G. Yang , C. Zhou, A. Reid, D. M. Hepburn, M. D. Judd and W. H. Siew, "Rough Set Theory Applied to Pattern Recognition of Partial Discharge in Noise Affected Cable Data," *IEEE Trans. Dielectr. Electr. Insul.,* vol. 24, no. 1, pp. 147-156, 2017.

[9] X. Peng, J. Li, G. Wang, Y. Wu, L. Li, Z. Li, A. A. Bhatti, C. Zhou, D. M. Hepburn, A. J. Reid, M. D. Judd and W. H. Siew, "Random Forest Based Optimal Feature Selection for Partial Discharge Pattern Recognition in HV Cables," *IEEE Trans. Pow. Del.,* vol. 34, no. 4, pp. 1715-1724, 2019.

[10] X. Peng, F. Yang, G. Wang, Y. Wu, L. Li, Z. Li, A. A. Bhatti, C. Zhou, D. M. Hepburn, A. J. Reid, M. D. Judd and W. H. Siew, "A Convolutional Neural Network-Based Deep Learning Methodology for Recognition of Partial Discharge Patterns from High-Voltage Cables," *IEEE Trans. Pow. Del.,* vol. 34, no. 4, pp. 1460-1469, 2019.

[11] V. M. Catterson and B. Sheng, "Deep Neural Networks for Understanding and Diagnosing Partial Discharge Data," in *2015 IEEE Electr. Insul. Conf. (EIC)*, Seattle, 2015.

[12] R. J. K. Wong, H. A. Illias and A. H. Abu Bakar, "High Noise Tolerance Feature Extraction for Partial Discharge Classification in XLPE Cable Joints," *IEEE Trans. Dielectr. Electr. Insul.,* vol. 24, no. 1, pp. 66-74, 2017.

[13] J. Yeo, H. Jin, C. Yuen and C. Ng, "Predicting of Partial Discharge in Medium-Voltage Cables Using Recurrent Neural Network with Long Short-Term Memory Cells after Wavelet Denoising," in *8th Int. Conf. Cond. Mon. Diag.*, Phuket, 2020.

[14] A. Rodrigo Mor, P. Morshuis, P. Llovera, V. Fuster and A. Quijano, "Localization Techniques of Partial Discharges at Cable Ends in Off-line Single-Sided Partial Discharge Cable Measurements," *IEEE Trans. Dielectr. Electr. Insul.,* vol. 23, no. 1, pp. 428 - 434, 2016.

[15] D. H. Hubel and T. N. Wiesel, "Receptive Fields and Functional Architecture of Monkey Striate Cortex," *The Jour. of Phys.,* vol. 195, no. 1, pp. 215-243, 1968.

[16] D. E. Rumelhart, G. E. Hinton and R. J. Williams, "Learning Representations by Back-Propagating Errors," *Nature,* vol. 323, pp. 533-536, 1986.

[17] S. Hochreiter, *Untersuchungen zu dynamischen neuronalen Netzen,* Munich: TUM.University Press, 1991.

[18] J. Yeo, H. Jin, C. Yuen and C. Ng, "Developing Neural network Models for Partial Discharge Analysis," in *46th Ann. Conf. IEEE Ind. Electr. Soc.*, Singapore, 2020.

[19] A. Rodrigo Mor, L. C. Castro Heredia and F. A. Munoz, "Estimation of Charge, Energy and Polarity of Noisy Partial Discharge Pulses," *IEEE Trans. Dielectr. Electr. Insul. ,* vol. 24, no. 4, pp. 2511-2521, 2017.

[20] H. Liao, "Speaker Adaptation of Context Dependent Deep Neural Networks," in *2013 IEEE Int. Conf. on Acous., Speech, Sign. Proc.*, Vancouver, 2013.

[21] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *3rd International Conference for Learning Representations*, San Diego, 2015.

[22] F. Chollet and others, *Keras,* https://keras.io, 2015.

[23] M. Abadi and others, *TensorFlow: Large-scale machine learning on heterogeneous systems,* tensorflow.org, 2015.