



21, rue d'Artois, F-75008 PARIS

<http://www.cigre.org>

CIGRE US National Committee
2021 Grid of the Future Symposium

Using Machine Learning and Deep Learning for Characterizing Partial Discharge in Underground Utility Cables for Predictive Maintenance Application

S. ZIEGLER
IMCORP
USA

S. SHEKHAR
MathWorks Inc.
USA

SUMMARY

Each year, millions of people and thousands of businesses are impacted by underground cable system failures. Underground cables are considered critical equipment within any power system, and typically one of the most expensive components of the system to repair. When they fail, the customer impact is immense and has the potential to cause severe collateral damage or worse, public safety concerns. Replacing underground power cables can be very expensive and time consuming and can take months or even years when associated with significant design, civil and construction work. According to IEEE, approximately 90% of all underground cable systems fail due to a breakdown of the solid dielectric insulation (i.e.: XLPE or EPR) ^[2].

Partial Discharge (PD) is an undesirable phenomenon that occurs when the strength of the electric field within a cable system exceeds the dielectric withstand capabilities of the insulation material. PD is both an insulation damaging process and a symptomatic signal of insulation breakdown ^[10]. Analyzing and interpreting results from PD measurements taken in the field is a complex task for humans (signal analysts) to perform. Interpretation of PD time domain signals (time-resolved partial discharge) depends on human decision making and can vary from person to person. With the emergence of Artificial Intelligence (AI) technologies, underground distribution cable system defects, which will lead eventually to failures, can be automatically predicted. Identifying cable defects before failure begins with the understanding of how cable system failures develop.

This paper characterizes the waveforms of Partial Discharge (PD) time domain signals utilizing a unique dataset ^[7] of measured conditions of underground power cable systems. Machine Learning and Deep Learning models have been developed and evaluated for the purposes of providing the foundation for automated condition monitoring and predictive maintenance. With the developments in Artificial Intelligence technologies, Deep Learning methods have been used for automated feature extraction and classification. This paper demonstrates that Deep Learning methods result in better accuracies than the typical Machine Learning (shallow learning) methods, providing more efficient automated identification techniques. The results demonstrate a step towards a predictive maintenance approach for underground cable systems.

KEYWORDS

Partial Discharge, Grid reliability, Asset Management, Predictive Maintenance, Extruded Dielectric Cables, Phase-Resolved Partial Discharge, Machine Learning, Deep Learning, LSTM, Time Series, Smart Diagnosis

1.0 INTRODUCTION

Efficient defect diagnosis is crucial to prevent unexpected and costly outages of electrical network systems. Each year, millions of people and thousands of businesses are impacted by underground cable system failures ^{[11] - [12]}. Most underground power cables installed since the early 1970s are solid dielectric cables (mainly Polyethylene-based and Ethylene-Propylene-Rubber-based) ^{[8], [10]}. Differentiating cables with “high to low-risk defects” along with those that are “defect-free” enables predictive cable system maintenance. Cable and accessory manufacturers have established a very reliable quality control standard and process in the past few decades that can be replicated by some partial discharge assessment technologies in the field ^{[13] - [15]}. These standards and processes are based on a 50/60 Hz sinusoidal voltage application (emulating operating conditions) and a measurement sensitivity baseline of 5 pC (pico-coulomb). These measurements are taken as the base for predictive maintenance.



Figure 1.0: Typical PD-producing defects in underground power cables (Top, left to right: installation defect, electrical tree defect, and manufacturing defect. Bottom: Failed cable - courtesy of General Cable)

Analyzing and interpreting results from partial discharge measurements taken in the field can be a complex task for humans ^[5]. Interpretation of PD signals depends on human decision making and can vary. It is time consuming, expensive, and difficult to scale. Also, different signal analysts may interpret PD signals differently, which can lead to inconsistencies in the quality of the assessment results. Nowadays, with the increasing presence of predictive analytics technologies, Smart Diagnosis implementation using more automated techniques must also be considered.

2.0 PARTIAL DISCHARGE BACKGROUND

Partial Discharge is an undesirable phenomenon that occurs when the strength of the electric field within a cable system exceeds the dielectric withstand capabilities of the insulation material. PD is both an insulation damaging process and a symptomatic signal of insulation breakdown. For a PD process to start, a defect must first exist in the extruded cable or the accessories, such as terminations and splices, which make up the cable system ^[16]. These defects can be introduced in several ways, either at the manufacturing process (seldom), during shipping and handling, installation craftsmanship, and lastly post installation damage. The ability to locate the origin of PD allows to predict where the asset will eventually fail.

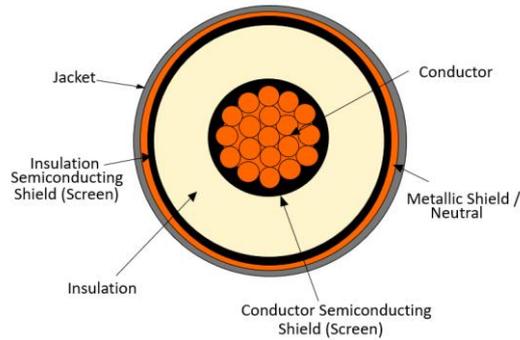


Figure 2.0: Typical construction of an underground power cable (cross-section)

To monitor and assess the electrical condition of underground power cables, partial discharge measurements are the most effective method. The location of a cable defect, caused by partial discharge, is determined by Time Domain Reflectometry (TDR) method, which uses the reflection of pulse signals at the cable terminals [6]. PD can be diagnosed online or offline. When choosing to measure with an online system (cable connected in service), it must be understood that partial discharge is unlikely to occur in the cable insulation, due to the lower electrical stress at operating condition. It is the transient overvoltage events that determine the development of PD defects in the cable insulation. In addition, if a PD defect was active at operating voltage, it would be active continuously and most likely fail fast with little time to react.

PD is measured with signal sensors. The output signals from the sensors can be represented in the time-domain as an individual PD pulse waveform as shown in Figure 3a. The phase-resolved PD (PRPD) pattern is widely used to characterize this phenomenon and represents the apparent charge (Q) versus the corresponding phase angle at which the individual PD events occur (ϕ) in the AC network voltage [9]. An example of PRPD is shown in Figure 3b.

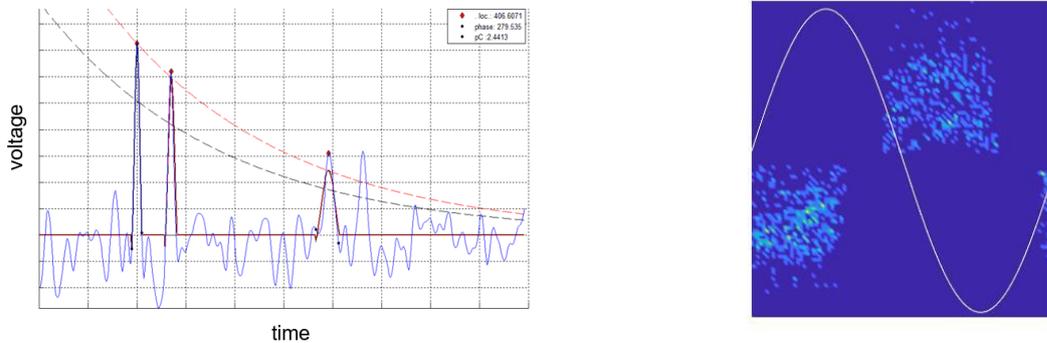


Fig 3.0: (a) Partial Discharge (PD) waveform in time-domain (b) Phase resolved Partial Discharge (PRPD) Pattern

There is a strong correlation between these phase-resolved partial discharge patterns and the type of PD sources (defects); therefore, a defect classification can be made by extracting discriminatory features from these representations and correlating them to known cases, usually obtained from root-cause analyses. Image processing techniques like texture analysis algorithms, fractal features, wavelet-based image decomposition can be applied to extract features from the PRPD image. Alternatively, signal processing techniques such as Fourier series analysis, Haar and Walsh transforms or wavelet transform, the features can be obtained from the PD time-domain signals. However, all these techniques provide semi-automatic feature extraction, and a lot of effort and expertise is required. Therefore, Machine Learning and Deep Learning techniques are used to automate feature extraction and classification of partial discharge signals. These two techniques are discussed in next sections.

3.0 MOTIVATION AND PURPOSE OF THIS STUDY

There are increasing initiatives globally for modernization of the grid infrastructure leading to the need for better condition monitoring and predictive maintenance of field equipment. Underground cables are critical grid infrastructure equipment that is less prone to severe weather events, such as windstorms, wildfires, ice storms, and lightning, compared to their overhead counterparts. But, when they fail, the customer impact is immense and has the potential to cause severe collateral damage or worse and public safety concerns. Underground fault detection and location practices are more complicated and time consuming than overhead power lines and typically mean longer outages and higher cost. Replacing underground cables can be very expensive and time consuming and can take months or even years when associated with significant design, civil and construction work. According to IEEE, approximately 90% of all underground cable systems fail due to a breakdown of the components insulation ^[2].

Analyzing and interpreting results from PD measurements taken in the field is a complex task for humans (signal analysts) to perform. Interpretation of PD time domain signals (time-resolved partial discharge) depends on human decision making and can vary from person to person. Different signal analysts may interpret signals differently, which can lead to inconsistencies in the quality of the results and in turn require additional human auditing to ensure quality. This process is time consuming, expensive, and difficult to scale.

Traditional analysis requires expert knowledge in interpreting PD time series signals. Expert knowledge in this field is difficult to establish or transfer through various teaching methods. Training times to become a reliable signal analyst can take months. Typically, user interfaces are provided that allow for different investigations and measurements on given PD time series, controlled by the user. It is a manual process that reviews PD time series signals in the order of occurrence. This process has proven very time consuming. In contrast, the new approach demonstrated in this study allows for the establishment of Deep Learning models to use the labels, generated by expert signal analysts, to auto-generate the best features on which the predictions are made. Deep Learning models are highly flexible and can be used at different points in the process of making predictions, including at the instrument hardware level.

The benefits of automating the characterization of signals obtained in PD measurements will result in a scalable process with known accuracy, while at the same time enabling cost saving opportunities coupled with resource task optimization.

This study intends to use predictive analytics methods ^{[1],[4]} for the following purposes:

1. **Classification** - An automated classification for the features and time series obtained from PD measurements in the field is performed to differentiate PD signals from ambient background noise signals.
2. **Location Identification** - A location range of origin is assigned to the positively identified PD signals.
3. **Defect Type Classification** - Insulation defects (causing PD signals to originate from the same defect location) are classified into different defect types by utilizing the Phase Resolved Partial Discharge (PRPD) representation.

4.0 PREDICTING CABLE SYSTEM FAULTS USING MACHINE LEARNING AND DEEP LEARNING METHODS

Success with AI depends on quality and quantity of data. An important factor to consider is “labelled data”. Information on asset types and defect types helps in better training of AI models and increases accuracy. However, labelled data is not always available as data labelling is often done by engineers and its very challenging to look at hundreds and thousands of signal waveforms and label them. This is where using unsupervised machine learning techniques can be very useful.

Unsupervised techniques allow for clustering and identification of data based on different auto-generated features and this could be a pre-step to help with the process of labelling the data if it’s not already available. To predict when and where power cable systems will fail, Partial Discharge measurements from underground cables were obtained.

4.1 FEATURE DATA

Partial discharge time series are digitized signals, picked up with a high-voltage and high-frequency coupler (sensor), which is connected to the cable-under-test [31]. The digitization is being performed by an Analog-to-Digital converter (ADC), sampling typically at 100 MHz and is based on a trigger mode. Therefore, the sample intervals are event-based and not time-based. Proprietary digital signal processing algorithms are used to extract features from the digitized signals for the purpose of characterizing these by “partial discharge” and “non-partial discharge”, PD location of origin and PD defect type.

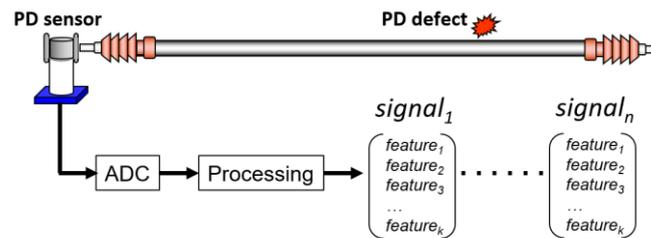


Fig 4.0: Capturing time series signals from an underground power cable

After converting signals into digital format and applying digital signal processing, each sampled event is being assigned a vector of features (F_i). The classification by feature vectors is based on trained machine learning or deep learning models.

In the first study, 4,076,749 labelled instances consisting of 23 features each, were provided. The feature data was obtained from a proprietary digital signal processing algorithm that processes the digitized time series signals. Out of the 23 features, 12 features were found to be representative and useful and were bundled into feature vectors. The labels (classes) are either “0” (non-PD) or “1” (PD).

4.2 TIME SERIES DATA

In the second study, 965,310 labelled and time-resolved partial discharge signals (digitized time series), with a class imbalance of 700,500 (non-PD) and 264,810 (PD) were provided. As these digitized time series were captured on underground power cables of different lengths, based on the time-domain-reflectometry principle, the time series are variable in length (sampling points).

The length of these sequences is between 119 and 1604 sampling points. To avoid excessive zero-padding, the data was subdivided into different length ranges.

In this approach time series signals were used directly as inputs to an LSTM (Long Short-Term Memory) model for classification. The studies were conducted in MATLAB. The binomial classification of signal

time series into “PD” and “non-PD” and the PD location assignment were done with and without data augmentation. In addition to the time series signal, the magnitude of the Fast Fourier Transform (FFT) or the magnitude of the Wavelet Transform (WT) was added as augmented data. The advantages of data augmentation were not significant. However, the computational requirements grew significantly by adding wavelet-transformed data.

5.0 METHODOLOGY

Conventional Machine Learning methods are sometimes preferred over Deep Learning if less data is available, or the computational requirements are an issue. Machine Learning algorithms are sometimes better to understand in terms of how the results are being produced.

Conventional (shallow) Machine Learning and deep learning methodologies were used in this study for PD classification. Conventional Machine Learning methods included Logistic Regression, Support Vector Machines, Random Forest Classifiers and Ensemble Learning (AdaBoost) classifiers.

6.0 RESULTS

Table 1 shows a summary of the results obtained from both, machine learning based and deep learning-based methodologies. The deep learning methodologies provided better results than the classic approaches. The deep learning approaches for this study include fully connected neural networks, recurrent neural networks, specifically LSTM networks, and convolutional neural networks.

Classifier	Accuracy	Precision	Recall	F1-score
Random Forest	0.90	0: 0.97 1: 0.70	0: 0.90 1: 0.90	0: 0.94 1: 0.79
Logistic Regression	0.91	0: 0.95 1: 0.74	0: 0.93 1: 0.82	0: 0.94 1: 0.78
Support Vector Machine	0.91	0: 0.95 1: 0.74	0: 0.93 1: 0.82	0: 0.94 1: 0.78
Ensemble Learning (AdaBoost)	0.91	0: 0.93 1: 0.83	0: 0.96 1: 0.71	0: 0.95 1: 0.77
Neural Network	0.95	0: 0.97 1: 0.86	0: 0.96 1: 0.87	0: 0.97 1: 0.87

Table 1.0: Binomial classification accuracies from shallow learning methodologies. Random Forest, Logistic Regression, Support Vector Machine using SGD and Ensemble Learning (AdaBoost). Neural Network (deep learning) classification accuracy listed in the last row

	precision	recall	f1-score	support
non-PD	0.98	0.96	0.97	631966
PD	0.84	0.91	0.87	158790
accuracy			0.95	790756
macro avg	0.91	0.93	0.92	790756
weighted avg	0.95	0.95	0.95	790756

Fig 5.0: Results from the binomial classification for numeric input features on the test set with a neural network approach

A workflow was developed for the LSTM approach of classifying digitized time series for the purpose of assigning these correctly into “PD” and “Non-PD” and assigning a location of origin for the “PD” class.

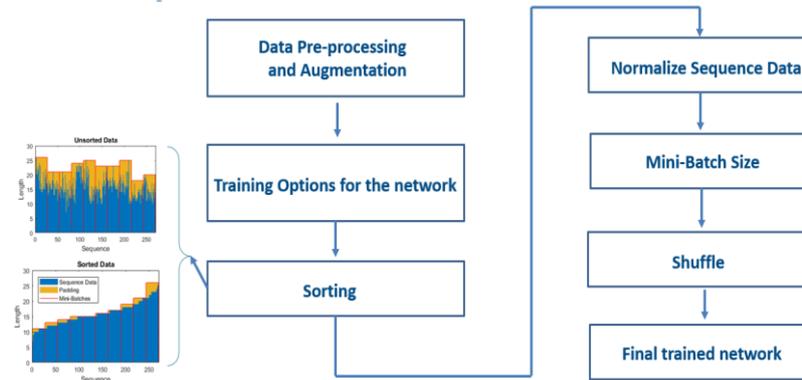


Fig 6.0: Deep Learning workflow for model training

The results for the PD classification into “PD” and “Non-PD” are displayed in figure 7. The classification, based on the digitized time series only, are plotted in blue color. It can be observed that for longer cable lengths, the accuracy of the supervised model slightly drops. This is due to the natural signal attenuation in cables and the diminishing clarity of the autogenerated features. The data augmentation efforts did not yield higher model accuracies. The FFT-augmented and wavelet transform-augmented data, respectively plotted in pink and black, demonstrate this.

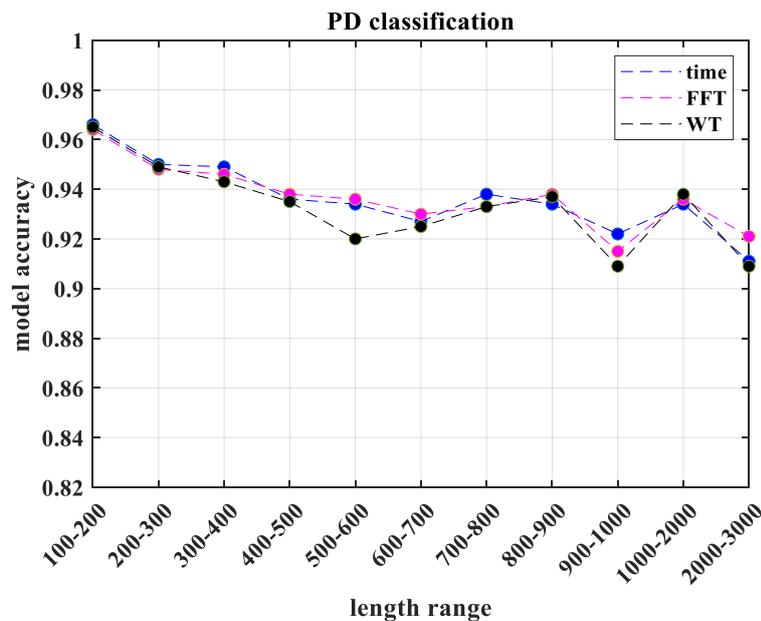


Fig 7.0: Binomial Classification results with and without data augmentation

The LSTM-based location assignment for positively identified PD signal waveforms (digitized time series) is based on a location range classification into 20 equally spaced classes. The example below is based on 281,989 time series signals. The test set results are as follows for the example of a location range from 200 ft. – 300 ft. The model accuracy on the test set was determined to be 93.27%.

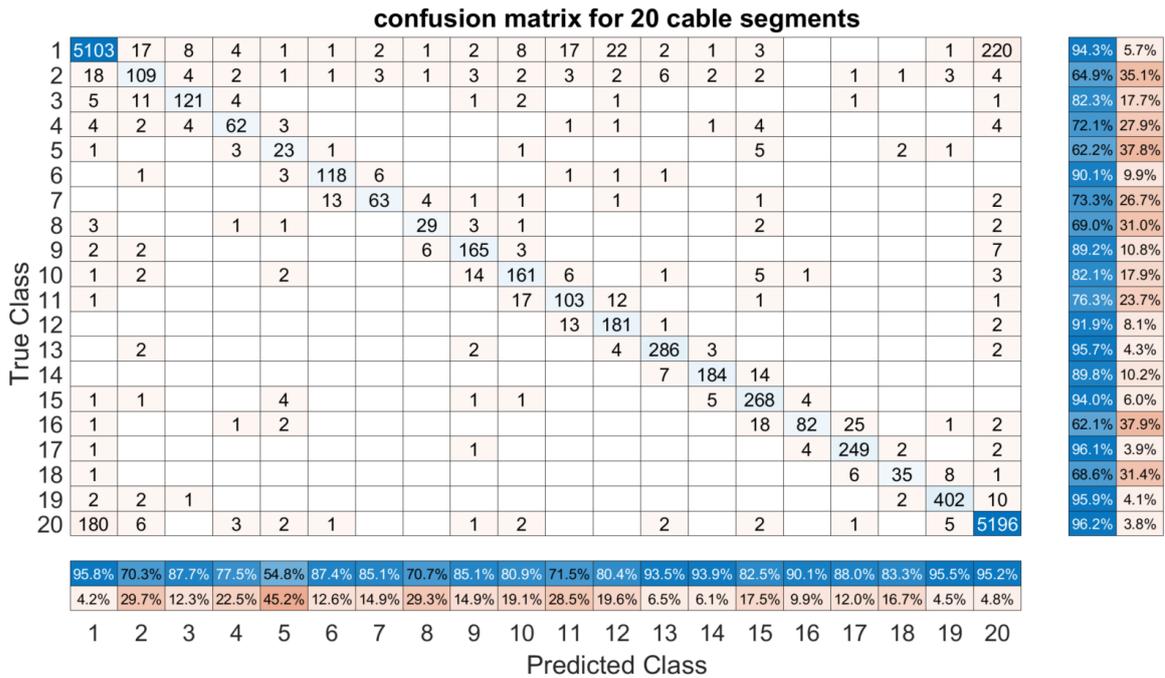


Fig 8.0: Confusion matrix results of positively identified PD into 20 location ranges on the test set

Deep learning workflows typically include convolutional neural networks (CNN) and recurrent neural networks (RNN), of which the LSTM approach is a specific case of an RNN. The latter is a good fit for ordered data (such as digitized time series). CNNs are very good with the task of image classification or object recognition within images.

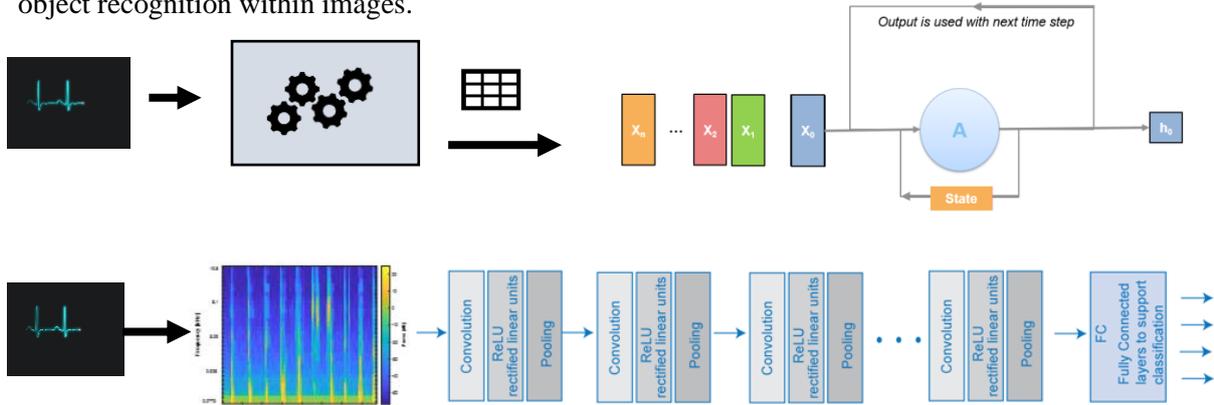


Fig 9.0: Deep Learning workflow for signal applications (top: LSTM, bottom: CNN)

Phase-resolved Partial Discharge (PRPD) plots are composed of positively identified partial discharge events occurring from one specific defect location. The phase-resolved PD (PRPD) pattern is widely used to characterize this phenomenon and represents the apparent charge (Q) versus the corresponding phase angle at which the individual PD events occur (ϕ). These plots are converted into coded images for automated defect type classification. Electrical Tree type defects are posing the highest risk of imminent cable system failure. Being able to determine if a PD is caused by an electrical tree or not allows assignment of a risk factor to a cable system asset. Paired with the measured partial discharge inception voltage (PDIV) of a defect, a weight to the risk factor is assigned. Hence, the risk categorization of assets is possible.

In the study below, 3,007 images with known defect types (5 different classes) were used as an input to train a CNN model. The defect types are usually determined from root-cause analyses of PD defects. The available data from an extensive data base [7] allowed for this classification.

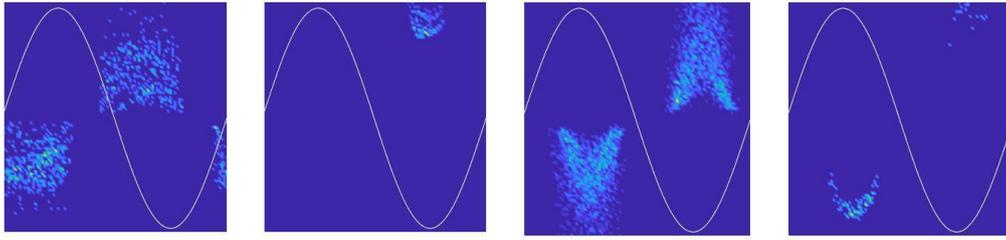


Fig 10.0: PRPD plots of different defect types (from left to right: termination PD, external PD, Electrical Tree, Electrical Tree)

The results for assigning the 3,007 PRPD images to 5 different defect classes are given below. The model accuracy was determined to be 96.7%.

class accuracy on test set: 96.7%

True Class	1	27	3		4	1	77.1%	22.9%
	2	1	21				95.5%	4.5%
	3	1		8			88.9%	11.1%
	4	1			250		99.6%	0.4%
	5					20	100.0%	

90.0%	87.5%	100.0%	98.4%	95.2%
10.0%	12.5%		1.6%	4.8%
1	2	3	4	5

Predicted Class

Fig 11.0: Confusion matrix results for PRPD image classification into 5 classes for the test set

7.0 LESSONS LEARNT AND KNOWLEDGE GAINED

Automated risk factor assignment to assets in underground cable systems allows for prioritization of repairs and ensuring that all materials, equipment, and human resources for the upcoming repairs are available. Strategic and proactive investment into the underground grid is necessary to ensure reliability and resilience. The presented methodology allows cable network owners and operators to base their strategic and proactive investments on actual health conditions of their underground cable systems.

8.0 CONCLUSION

The models introduced in this report describe a methodology to successfully achieve a workflow for risk assignment to partial discharge-based defects in underground power cable systems. The deep learning signal waveform characterization of the PD discussed in this paper now introduces a predictive maintenance component to the underground asset management.

BIBLIOGRAPHY

- [1] Aurélien Géron - Hands-on Machine Learning with Scikit-Learn, Keras and TensorFlow
- [2] 493-1997 - IEEE Recommended Practice for the Design of Reliable Industrial and Commercial Power Systems (Gold Book).
- [3] Partial Discharge location as a diagnostic tool for power cables. M.S. Mashikian. 2000 IEEE Power Engineering Society Winter Meeting. Conference Proceedings (Cat. No.00CH37077)
- [4] M. Kudo, J. Toyama and M. Shimbo. (1999). "Multidimensional Curve Classification Using Passing-Through Regions". Pattern Recognition Letters, Vol. 20, No. 11--13, pages 1103—1111
- [5] IEEE 400 Guide for Field Testing and Evaluation of Insulation of Shielded Power Cable Systems, Piscataway, NJ: IEEE.
- [6] M.S. Mashikian, "Preventive Maintenance Testing of Shielded Power Cable Systems", IEEE Transactions on Industry Applications, Vol. 38, No. 3 May/June '02.
- [7] The IMCORP database contains more than 76 million meters (250 million feet) of shielded power cable systems, ranging from 5kV to 500kV that were tested in 15 countries. The database includes a meter-by-meter condition profile of over 210,000 cable phases, including the underlying digitized signals. [status: 11/2020]
- [8] Essay Wen Shu, Nigel Hampton, "Evolution of MV extruded Cable Designs Used in the US from 1996 to 2014", Jicable 2019
- [9] Sonia Barrios, David Buldain, Maria Paz Comech, Ian Gilbert, Inaki Orue - "Partial Discharge Classification Using Deep Learning Methods - Survey of Recent Progress"
- [10] Philip Cox "EPR, XLPE, Age vs. ACBD" - IEEE PES ICC Spring 2015 Sub A Minutes A290 of A373 Slide 23
- [11] U.S. Energy Information Administration <https://www.eia.gov/todayinenergy/detail.php?id=35652>
- [12] U.S. Energy Information Administration <https://www.eia.gov/todayinenergy/detail.php?id=43915>
- [13] IEEE 48-2020 - IEEE Standard for Test Procedures and Requirements for Alternating-Current Cable Terminations Used on Shielded Cables Having Laminated Insulation Rated 2.5 kV through 765 kV or Extruded Insulation Rated 2.5 kV through 500 kV
- [14] IEEE 404-2012 - IEEE Standard for Extruded and Laminated Dielectric Shielded Cable Joints Rated 2.5 kV to 500 kV
- [15] IEEE 386-2016 - IEEE Standard for Separable Insulated Connector Systems for Power Distribution Systems Rated 2.5 kV through 35 kV
- [16] Mashikian M., Szatkowski A., "Medium Voltage Cable Defects Revealed by Off-Line Partial Discharge Testing at Power Frequency", IEEE Electrical Insulation Magazine, DEIS Feature Article, Vol. 22, No. 4 July/August 2006