Intelligent Monitoring of Transformer Insulation using Convolutional Neural Networks

Wei Lee Woon[1], Zeyar Aung[1], and Ayman El-Hag[2]

[1] Department of Computer Science, Masdar Institute
(Khalifa University of Science and Technology),
P.O. Box 127788, Abu Dhabi, UAE.
[2] Department of Electrical and Computer Engineering,
University of Waterloo, 200 University Avenue West,
Waterloo, Ontario,Canada, N2l 3G1.
{wei.woon, zeyar.aung}@ku.ac.ae ; ahalhaj@uwaterloo.ca

Abstract. The ability to monitor and detect potential faults in smart grid system components is extremely valuable. In this paper, we demonstrate the use of machine learning techniques for condition monitoring in power transformers. Our objective is to classify the three different types of Partial Discharge (PD), the identify of which is highly correlated with insulation failure. Measurements from Acoustic Emission (AE) sensors are used as input data. Two broad machine learning based approaches are considered - the conventional method which uses a predefined feature set (Fourier based), and deep learning where features are learned automatically from the data. The performance of deep learning compares very favorably to the traditional approach, which includes ensemble learning and support vector machines, while eliminating the need for explicit feature extraction from the input AE signals. The results are particularly encouraging as manual feature extraction is a subjective process that may require significant redesign when confronted with new operating conditions and data types. In contrast, the ability to automatically learn feature sets from the raw input data (AE signals) promises better generalization with minimal human intervention.

Keywords: Machine learning, Smart Grid, Deep Learning, Power Transformer, Convolutional Neural Networks, Partial Discharge

1 Introduction

1.1 Background and Motivation

Power transformers are amongst the most expensive and important assets in transmission and distribution systems, and can remain in service for over a century [1, 2]. Failure in a typical 100 MVA power transformer is a very serious event and can result in severe disruptions to business activities and energy efficiency initiatives, which could have serious environmental and financial consequences.

The aging of the transformer insulation system during its operational life is a natural phenomenon as transformers are constantly subjected to a range of electrical, mechanical, and thermal stresses [3]. Degradation in the insulation system frequently manifests as a localized dielectric breakdown known as a Partial Discharge (PD) [4]. Hence, detecting and assessing PDs can help prevent more serious transformer failure. Unfortunately, conventional monitoring techniques require the physical presence of a site engineer to diagnose problems. More recently, machine learning techniques [5] have been proposed as a way of automatically detecting PDs in transformer insulation, which could be significantly more cost-effective and reliable.

PDs can be detected in a number of ways. One interesting option is via the use of acoustic emission (AE) sensors[4], which are noninvasive, cost effective and easily installed by magnetically attaching the sensors to the transformer's tank wall - a procedure that can be undertaken even when the transformer is energized.

Furthermore, PDs can originate from different sources, each of which is indicative of different fault types and severities. As such, in this paper, we will focus on the *classification* of different PD types using machine learning methods to analyze data generated by AE sensors.

1.2 Novelty and Objectives

There have been prior attempts at using AE signals to identify PDs (see for e.g. [1, 6]). In [7], a wide band piezoelectric transducer, DC-1 MHz, was used to measure eight simulated PD types. An artificial neural network was used to classify Power Spectral Density (PSD) and Short Time Fourier Transform (STFT) features, resulting in recognition rates of over 90%. However, measurement conditions such as the oil temperature were not considered. In more recent work [6], AE sensors were used to study the effect of increasing the tank size, the presence of barriers in the insulating medium and oil age on PD detection capability. High recognition rates (96-100%) were obtained using spectral and statistical features, even when barriers were placed between the PD source and the AE sensor. However, the recognition rate dropped to 60-88% when old insulation oil was used.

In previous studies conducted by the authors of this paper, the classification of PDs initiated under different experimental conditions were presented [5, 8]. In agreement with other prior work, when the training and test datasets consisted of samples collected under similar conditions, high PD classification rates were achieved. However, recognition rates were lower when test sets collected under different experimental conditions were used.

In this study, our aim is to study the use of deep learning techniques on this challenging problem. These methods have been used with great success to analyze natural signals such as sound waves and digital images, and it would be very interesting to gauge the applicability of deep learning methods in the present context, and to characterize the performance characteristics *vis-a-vis* conventional methods.

2 Methods and Data

2.1 Classification

Classifications algorithms intelligently discriminate between objects or instances from different classes by identifying a suitable decision boundary (or boundaries) in the feature space in which the instances are embedded. A variety of different algorithms have been devised, with significant differences in performance characteristics.

As benchmarks, we use the methods presented in [5], which are the Random Forest, Gradient Boosting, SVM, Decision Tree and LDA algorithms. A detailed explanation of each of these algorithms is beyond the scope of the paper and the interested reader is referred to [5] as well as the many excellent references in the literature, for e.g. [9, 10].

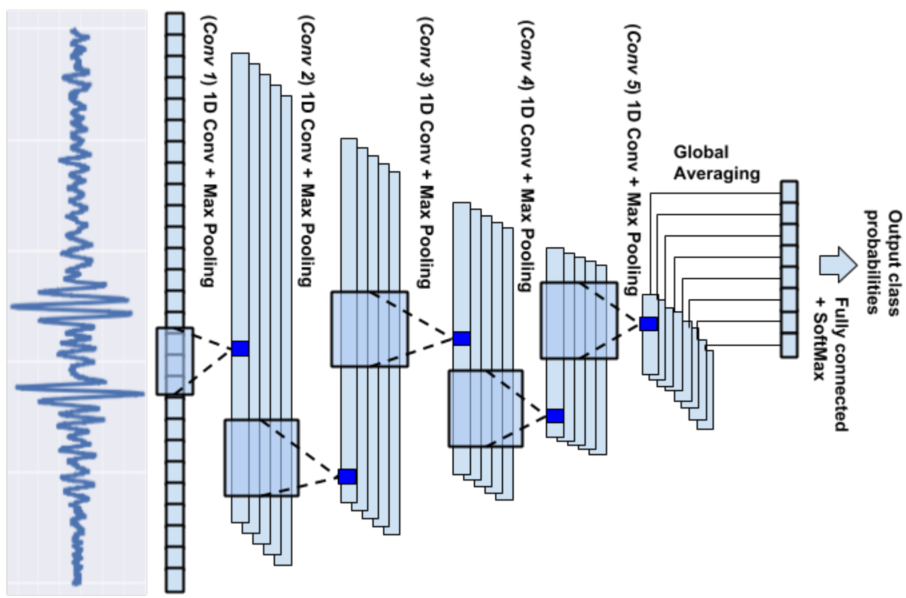

Fig. 1. 1D-CNN - Model architecture

2.2 1D-Convolution Neural Networks

A Convolutional Neural Network (CNN, or ConvNet) is a type of deep learning architecture that is commonly used in computer vision, but which can also be used for other forms of signal processing. CNNs are feedforward neural networks which contain one or more convolutional layers; these are composed of specialized units or *kernels* which filter inputs from a finite subset of the input data.

Traditional neural networks are “fully connected”, i.e. all the nodes in one layer are connected to all of the nodes in the next - this has two big disadvantages (i) it greatly increases the number of trainable parameters and, as a consequence, the amount of training data that is required (ii) when processing signals, the presence of a feature (say, a spike in a time series) may be important but not the exact position of the feature in the time series. A fully connected network will treat such occurrences as entirely different phenomena.

CNNs do not suffer from the above shortcomings, and are able to learn customized features automatically from the data itself, as opposed to the use of manually selected features, such as Fourier coefficients. Many of the notable applications of CNNs have been in machine vision, but 1-dimensional CNNs can equally be applied to time series and sequence data. Motivated by the advantages mentioned above, this is the approach adopted in this study.

The architecture used in this study is shown in Fig. 1. The configurations for each convolutional layer were determined using a randomized parameter search, and are shown in Tab. 1.

Table 1. Configurations for convolutional layers

Layer	Location 4	Location 9	Unheated oil	Heated oil
Conv 1	100@10 $\times$ 1	100@10 $\times$ 1	110@9 $\times$ 1	89@8 $\times$ 1
Conv 2	150@10 $\times$ 1	150@10 $\times$ 1	165@9 $\times$ 1	133@8 $\times$ 1
Conv 3	150@10 $\times$ 1	150@10 $\times$ 1	165@9 $\times$ 1	133@8 $\times$ 1
Conv 4	150@10 $\times$ 1	150@10 $\times$ 1	165@9 $\times$ 1	133@8 $\times$ 1
Conv 5	200@10 $\times$ 1	200@10 $\times$ 1	220@9 $\times$ 1	178@8 $\times$ 1

After each convolution layer, 3×1 Max-Pooling was applied, while for the final layer we used Global Average Pooling (GAP) followed by Softmax. GAP was used to reduce the total number of parameters in the model, an important consideration given the small size of the training data sets.

2.3 Experimental Setup

Three different PD classes were generated, namely: (i) Discharges from a sharp point to ground plane (ii) Surface discharges (iii) Discharges from a void in the insulation. A $1 \times 1 \times 0.5$m tank filled with aged oil received from a local utility company was used to conduct the experiments.

An AE sensor with a resonance frequency of 150kHz and bandwidth of 100-450 kHz was used, while data acquisition was performed using an oscilloscope interfaced with Matlab™, with the sampling frequency set to 10M sample/sec for a window of 2500 samples (250 μsec). A magnetic holder is used to attach the AE sensor to the tank wall and silicone grease was applied to improve acoustic coupling.

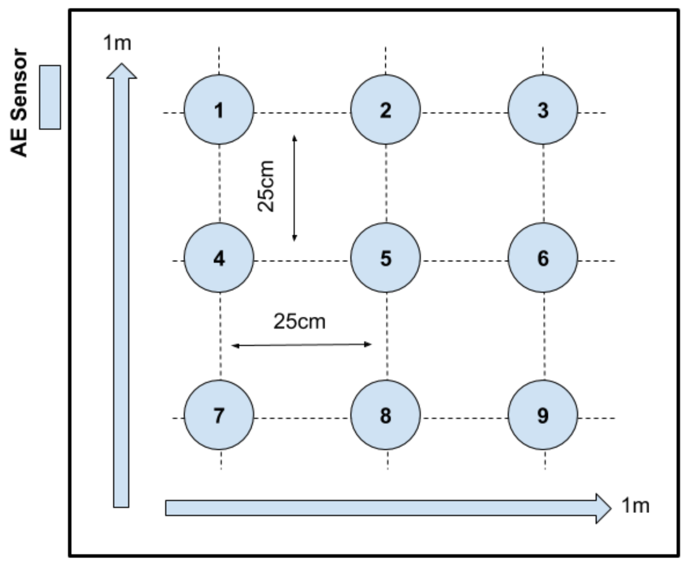

Fig. 2. Location of PDs relative to AE sensor [5]

2.4 Measurement Conditions

As mentioned previously, the more challenging classification problem is when data from the different PD types were collected under variations in the measurement conditions. In this paper we focus on two such variations:

- **Location of the PD Source** PDs can be initiated at any physical point in the insulation system. Fig. 2 shows the layout of the measurement tank annotated with the location of the AE sensor. As can be seen, the tank contains nine holes which allow the PD sources to be generated at different locations. For the results in this paper, the different PD types were simulated at locations 4 and 9.
- **Temperature of Insulating Oil** The temperature of a power transformer's oil varies according to operating conditions, and can be as high as 80^oC. This can affect the classifier performance as AE waves in oil have different propagation speeds at different oil temperatures. Tab. 2 shows the effect of increases in the oil temperature.

Table 2. Speed of AE waves at different oil temperatures [4]

Oil Temperature (^{o}C)	AE wave speed (m/s)
20	1413
50	1300
80	1200
110	1100

Note that in [5], the presence of a barrier in the insulation medium was also studied but the corresponding data is less complete and this was excluded from this paper due to space limitations. The total number of measurements for each of the PD types and measurement conditions are shown in Tab. 3

Table 3. Count of PD instances measured for each PD type and measurement condition

PD Type	Sharp	Surface	Void
Location #4	116	253	253
Location #9	143	233	472
Unheated oil	0	393	473
Heated oil	0	216	216

2.5 Data Pre-Processing and Analysis

Computational Tools All the development and data processing in this paper was performed using the Python programming language. Python has a concise, intuitive syntax which allows rapid development of a variety of different applications and tools. It also has an extremely wide variety of toolboxes which can support most scientific and technical computing tasks. Classification algorithms were implemented using *Keras*[3] for the CNNs, and *scikit-learn*[4], a commonly used and flexible Python machine learning toolkit.

Feature Extraction While CNNs are capable of learning customized feature sets, the benchmark classification algorithms require a more traditional feature

[3] https://keras.io/
[4] http://scikit-learn.org

extraction process, where instances to be classified are first converted into a feature vector which adequately characterizes these instances. As per [5], we used a Discrete Fourier Transform (DFT) based method to convert each time series into PSD histograms. The PSDs were extracted from the time series using the *periodogram* function from the well known *SciPy*[5] Python toolkit.

Validation Two numerical metrics were used to validate the performance of the classification algorithms:

1. **Accuracy**: the number of instances correctly classified as a percentage of the total number of instances in the test set. Accuracy is a reasonable metric to use where the number of training instances in the different classes are fairly well balanced, as is the case in this study.
2. **AUC**: The **A**rea **U**nder the **C**urve of the Receiver Operating Characteristic (ROC) curve for a given classifier. The ROC curve is the plot of true *vs* false positives, and allows the quality of a classifier to be evaluated. While accuracy only evaluates the operation of the classifier at a particular decision threshold, ROC rates overall performance and can help to detect cases where a particular classifier may be highly accurate, but is also very sensitive to minor shifts in the data or parameters.

Finally, for the cases where classifiers are trained and tested using data from the same recording session, cross-validation is used in combination with these numerical scores to ensure that a fair evaluation of the performance is obtained. More detailed discussions of these evaluation techniques can be obtained from [9].

3 Results

The results produced after conducting the experiments described earlier will now be presented and discussed.

3.1 Effect of Measurement Location

The first set of experiments targeted the effect of PD source location relative to the accoustic sensors. The results of these experiments are presented in Table 4. Some general observations:

1. The classification performance for experiments involving only individual locations (as measured using cross validation) were very high. Better scores were obtained in the case of Location 9.
2. When only a single measurement location at a time was considered (cross validation), the highest accuracies were obtained using the SVM classifier and Random Forest, but the CNN classifier was always very close.

[5] http://www.scipy.org

Table 4. Classification accuracies and AUC scores (different sensor locations). Highest scores per column are shown in bold.

Accuracy	4 (CV)	9 (CV)	4→9	9→4
CNN	88.8%	95.2%	**85.5%**	71.9%
Random Forest	85.2%	**99.6%**	80.0%	**73.2%**
Gradient Boosting	84.7%	99.1%	75.0%	71.4%
SVM	**89.3%**	97.1%	75.4%	72.1%
Decision Tree	81.4%	96.2%	70.2%	59.8%
LDA	79.6%	95.5%	72.9%	63.9%
AUC scores				
CNN	**0.94**	0.99	0.76	0.70
RandomForest	0.89	**1.00**	**0.87**	**0.83**
GradientBoosting	0.91	**1.00**	0.83	0.78
SVM	0.93	**1.00**	0.71	0.76
DecisionTree	0.82	0.95	0.62	0.51
LDA	0.81	0.97	0.58	0.61

3. However, when the classifiers were trained and tested using data collected at different locations, there was a sharp drop in accuracy for all classifiers. However, the CNN classifier produced the highest accuracy by a significant margin when classifying partial discharges at Location 9, and was very close to the best classifer (Random Forest) for Location 4.
4. The pattern in the AUC scores is similar to accuracy, but CNN obtained the top score for the cross validation case with Location 4.

3.2 Effect of Oil Temperature

The experiments were repeated at a single fixed location, but at different oil temperatures. Results are presented in Tab. 5. Some observations:

1. As with the previous set of results, the cross validation results were high across all the classifiers.
2. For the cross-condition experiments, the CNN classifier had the highest score when classifying partial discharges in heated oil.
3. However, the performance on the unheated oil was a little disappointing as the Random Forest and Gradient Boosting techniques both scored significantly higher, though CNN still outperformed the methods used in [8] (SVM,

Table 5. Classification accuracies and AUC scores (different oil temperatures). Highest scores per column are shown in bold.

Accuracy	Cold (CV)	Hot (CV)	Cold→Hot	Hot→Cold
CNN	97.7%	94.1%	**97.1%**	77.1%
RandomForest	99.2%	**99.7%**	89.8%	84.6%
GradientBoosting	**99.4%**	97.4%	92.6%	**85.7%**
SVM	97.8%	94.9%	50.0%	51.7%
DecisionTree	96.2%	97.5%	76.2%	72.6%
LDA	96.9%	94.8%	67.8%	57.5%
AUC scores				
CNN	0.99	0.97	**0.99**	0.88
RandomForest	**1.00**	**1.00**	**0.99**	**0.99**
GradientBoosting	**1.00**	**1.00**	0.96	0.95
SVM	0.99	0.99	0.42	0.70
DecisionTree	0.96	0.98	0.76	0.74
LDA	0.99	0.96	0.65	0.68

Decision Tree and LDA). However this disparity is not surprising as there were far more training samples for the unheated oil (on which the classifier for the heated oil used here is trained).

4 Discussions

Overall, the results of the experiments were promising and strongly motivate the need for future studies on the use of deep learning to identify faults in power transformers.

While the results obtained using convolutional neural networks were not a drastic improvement over Random Forest and Gradient boosting they were superior in two of the four cross-condition experiments (which are more important) and were very comparable in the other cases.

Note that this is actually an impressive result as there was unfortunately only very little data available, whereas CNNs typically require much larger training sets. In contrast, ensemble classifiers are considered top of the range amongst the "conventional" machine learning techniques and would be expected to produce very good performance in such circumstances. This view is substantiated by the observation that CNNs performed best in the cases where more data was available.

Finally, the results support the notion that AE sensors could be very valuable for intelligently monitoring the condition of power transformers. These findings could have important implications for the design of automatic power transformer monitoring schemes. In the future, it is hoped that innovations based on these results could be developed and tested, e.g. for deployment in predictive maintenance systems.

References

1. Harbaji, M., El-Hag, A., Shaban, K.: Accurate partial discharge classification from acoustic emission signals. In: Electric Power and Energy Conversion Systems (EPECS), 2013 3rd International Conference on. pp. 1–4. IEEE (2013)
2. Brown, M.H., Sedano, R.P.: Electricity transmission: a primer. Natl Conference of State (2004)
3. Kuo, C.C., Shieh, H.L.: Artificial classification system of aging period based on insulation status of transformers. In: Machine Learning and Cybernetics, 2009 International Conference on. vol. 6, pp. 3310–3315. IEEE (2009)
4. Sikorski, W., Ziomek, W.: Detection, recognition and location of partial discharge sources using acoustic emission method. Acoustic Emission pp. 49–74 (2012)
5. Woon, W.L., El-Hag, A., Harbaji, M.: Machine learning techniques for robust classification of partial discharges in oilpaper insulation systems. IET Science, Measurement & Technology 10(3), 221–227 (2016)
6. Swedan, A., El-Hag, A., Assaleh, K.: Acoustic detection of partial discharge using signal processing and pattern recognition techniques. Insight-Non-Destructive Testing and Condition Monitoring 54(12), 667–672 (2012)
7. Boczar, T., Borucki, S., Cichon, A., Zmarzly, D.: Application possibilities of artificial neural networks for recognizing partial discharges measured by the acoustic emission method. Dielectrics and Electrical Insulation, IEEE Transactions on 16(1), 214–223 (2009)
8. Harbaji, M., El-Hag, A., Shaban, K.: Classification of common partial discharge types in oil-paper insulation system using acoustic signals. IEEE Transactions on Dielectrics and Electrical Insulation (2015), accepted for publication
9. Hand, D.J., Mannila, H., Smyth, P.: Principles of data mining. MIT press (2001)
10. Hastie, T., Tibshirani, R., Friedman, J., Hastie, T., Friedman, J., Tibshirani, R.: The elements of statistical learning, vol. 2. Springer (2009)