Localization of Partial Discharge in Power Transformer Winding Using Sparse Autoencoder

H. MORADI TAVASANI, S. MANTACH, M. GUNAWARDANA, B. TABEI, and B. KORDI
University of Manitoba
Winnipeg, MB, Canada R3T5V6

SUMMARY

Transformers are the most expensive single pieces of equipment of an electric power system, and their failure considerably impacts power network performance. Partial discharges (PD) are local electrical discharges that might happen inside the insulation of transformers. Prolonged exposure to PD has detrimental effects on the integrity of a transformer's insulation system. As such, the detection and localization of PDs have been considered as one of the monitoring methods to ensure the reliable operation of transformers. Localization of PDs in transformer winding is more complicated than the detection of the occurrence of PDs. Machine learning techniques have been employed to analyze the PD signals, but the feature extraction process is always challenging. These techniques require the experience of experts and yet, in some cases, are not able to determine the exact location of PD. In this paper, to automate feature extraction, a combination of wavelet transform, a sparse autoencoder (SAE) and a logistic regression, one-vs-all classifier was employed to localize PD in the transformer winding. Using the SAE, no feature extraction analysis is required, which allows more accurate detection of PD location in transformer windings. A detailed high-frequency winding model is needed to construct a robust classification model. This winding model should properly simulate PD signal propagation in transformer windings. The axial multiconductor transmission line (AMTL) was used for the simulation of transformer winding that takes the variation of the turn radius into account. A continuous winding with several sections, each consisting of a number of turns, was simulated. The middle turn was selected at each section to inject the PD pulse signal. The PD signals were simulated by Gaussian pulses with a fixed peak and variable rise time. The injected signals propagated through the winding and were recorded at the grounded end of the winding. The PD pulse current measured at the ground terminal was used as raw data for the proposed classifier model. After hyperparameter tuning, the model could localize the PD signals with 99.1% training and 97.9% test accuracy.

KEYWORDS

Transformer winding, partial discharge, partial discharge localization, deep learning, machine learning, autoencoder, wavelet transform.

Corresponding author email: behzad.kordi@umanitoba.ca

1. INTRODUCTION

A typical transformer is comprised of two main windings and a magnetic core. Winding failure is the main cause of power transformer disruptions [1]. In the study of failures in transformer winding, partial discharge (PD) localization has been an increasingly important area. One of the main obstacles for PD localization is the lack of easy access to the internal turns of windings. Hence, finding a correlation between a measured PD pulse (for example, at the ground termination of the winding) and its location in the winding plays an essential role in addressing the issue of transformer failure. Attempts have been made to detect PD within transformers, including acoustic, ultra-high frequency (UHF), and electrical methods [2]. In [3], the winding transfer function has been used for the localization of PD. Since this method is sensitive to noise, the performance can be significantly influenced in a noisy medium. Reference [4] presents a Kullback–Leibler divergence method that calculates the divergence between the PD signal and the reference signal of each location. The least divergence value determines the location of PD. Another approach is based on calculating the correlation coefficients between the PD signals and reference signals [5]. The dependency of these methods to reference signals prevents a fast response, which is not desired for online applications. Some studies, such as [6], suggest condition monitoring of transformers due to faults in the winding by using thermal images acquired from the exterior of the transformer, which is not practical for PD localization.

This research proposes a simulation model to tackle the problem of PD localization. The first step involves a propagation model to generate a PD pulse dataset. Since the nature of PD signals is associated with fast-rising pulses, a high-frequency distributed model is required to accurately simulate PD pulse propagation along transformer winding [7]. Previous studies demonstrated that among the proposed winding models, the multiconductor transmission line (MTL) approach offers the widest bandwidth [7], [8]. A modification of the MTL-based models, the so-called axial multiconductor transmission line (AMTL) model, has been proposed in [9], where the variation of winding radius in each section is considered. Given its accuracy, this is the model employed for PD pulse propagation analysis in transformer winding in the present research.

The next step for PD localization is to analyze the propagation of PD pulses along the transformer winding. Several machine learning (ML) techniques have been employed for this purpose [7], [11]. These techniques sometimes suffer from overfitting and cannot be generalized to other transformer types, which results in repeating procedures. Hence, the feature extraction process is always a serious challenge faced by ML algorithms. The distribution of energy values has been used in [10] as PD signal features, and in [12], the permutation entropy of the PD signals is proposed as a feature. Four types of features were extracted and compared in [13], including statistical parameters, statistical parameters from wavelet transform, energy parameters from wavelet transform, and cross wavelet transform (XWT) parameters. These methods cannot be conducted without experienced experts and, in most cases, are not able to find the exact PD location. Deep learning (DL) techniques can overcome this issue by eliminating the need for hand-crafted feature extraction. However, the main drawback of such approaches is that they require a large dataset which is not feasible in all transformers. Furthermore, online and real-time monitorings are inconvenient due to their computational complexity. Finally, they are prone to overfitting because of a high number of hyperparameters in the model.

In this study, the hand-crafted features are substituted by those extracted using a sparse autoencoder (SAE). Furthermore, a combination of the common types of features available in the literature is investigated, and the results are compared with those by the SAE. The structure of the paper is as follows. Details of transformer winding modelling using AMTL is discussed in Section 2.1. The procedure of generating PD signals is proposed using a Gaussian waveform in Section 2.2. The approach of collecting the PD dataset and feature analysis is given in Sections 2.3 and 3, respectively. Wavelet transform is introduced as a tool to improve the feature analysis. Different scenarios for extracting the features from the wavelet transform of the PD signals are investigated. Finally, a logistic regression, one-vs-all classifier is employed to classify PD signals. That is to determine the location of PD in the winding.

2. PD PULSE PROPAGATION SIMULATION

2.1. Transformer Winding Model

In this work, an eight-disk winding with 128 turns (16 turns in each disk) was selected for modelling using the AMTL approach [9] (see Figure 1). Employing the AMTL for transient analysis of transformer winding, each turn was considered as a conductor of the multiconductor transmission line. In a typical MTL formulation, the main assumption is that the differences in the turns' length are negligible, so all transmission lines have the same length. However, in the AMTL approach, the transformer winding is modelled in a cylindrical coordinate system, and the length of the winding turns is not constant anymore. The first step of developing the AMTL simulation model was to determine the per-unit-length (PUL) parameters, i.e. the PUL inductance, capacitance, resistance, and conductance matrices. Each PUL parameter was represented as an $n \times n$ matrix where n was the number of turns ($n = 128$ in this work). To calculate the PUL capacitance and inductance matrices, finite element method (FEM) solution of electrostatic and magnetostatic problems was employed using a commercial software. Since this paper deals with high-frequency signals, the transformer core and tank can be assumed perfect electrical conductors (PEC) with high accuracy [14]. For each of the eight sections, the middle turn was selected for the injection of the PD signal, i.e. the possible locations for the PD source would be turns 8, 24, 40, …, 120, which were labelled as classes 1, 2, 3, …, 8. The PD signal source was modelled as an ideal current source with a Gaussian waveform (see Section 2.2). The end of the last turn of the transformer winding was grounded via a small resistance, while the other end of the winding was open circuit.

2.2. PD Pulse Simulation

Each PD signal was represented by a single Gaussian pulse with a specified peak and rise time [15]. The rise time can be influenced by the PD source and the insulation type. However, the PD pulse is typically measured at a location far from its source. In this work, the peak of the injected PD pulse was considered to be constant, and variability of the data for each class was achieved by varying the rise time in the range of 0.5 ns to 10 ns. The Gaussian pulse is defined as

$$i(t) = Ae^{-(t-t_0)^2/\alpha^2} \tag{1}$$

where A is the magnitude of the current and t_0 and α determine the centre time and the rise time of the injected PD pulses. Heidler's function can also be a good representation of PD signals since, unlike the Gaussian function, it has different values for the rise time and the fall time [16].

2.3. PD Dataset

The Gaussian pulse was injected into the locations identified in Section 2.1, assuming 20 different rise times, starting from 0.5 ns to 10 ns with a step of 0.5 ns. An amplitude of unity was assumed for the injected PD pulse in the simulation. Changing the rise time as well as the location of the PD source provides the variability for our classification model. To simulate the PD signal measured at the ground connection of the winding, finite-difference time-domain (FDTD) analysis was employed. More details on the FDTD technique used for the simulation of the MTL system can be found in [17]. The simulation was conducted with a time step of 0.3 ns. The current of the last turn, the grounded turn, was stored in each time iteration as an output.

3. FEATURE ANALYSIS

3.1. Wavelet Transform

PD signals contain high-frequency components associated with a typical rise time of a few nanoseconds. This implies that the high-frequency components of the measured PD signals contain effective information about the PD source. In such cases, the wavelet transform of the signal can improve the interpretability of the data as it maps time-domain PD signals into two groups of coefficients. Unlike the Fourier Transform that provides frequency-domain information only, the wavelet transform represents the PD signal in both time and frequency domains. In other words, using the wavelet transform, not only the frequencies of a signal but also the time of occurring can be determined.

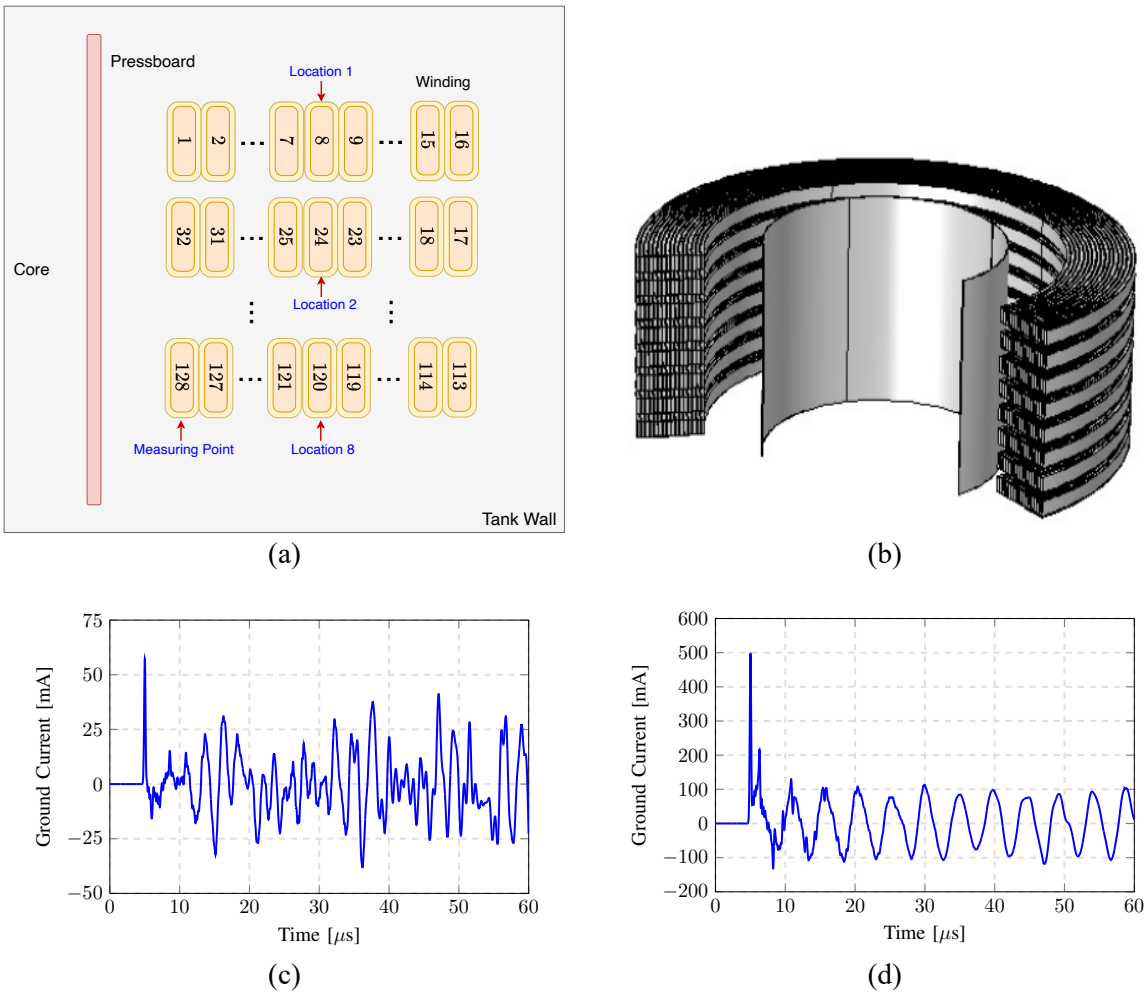

Figure 1. (a) Schematic of 2D cross section of the winding and (b) and 3D split view. Two sample current waveforms recorded at the ground terminal when the PD pulse is injected at (c) location 1 and (d) location 8.

Continuous wavelet transform (CWT) and discrete wavelet transform (DWT) are the two types of wavelet transform. In this work, DWT is used as it is computationally less expensive. The other drawback of CWT is producing redundant data. DWT can be considered as a filter bank that applies a high-pass and low-pass filter to the signal at the so-called decomposition levels and represent the signal as two arrays of approximation and detail coefficients. The approximation and detail coefficients are associated with the low-pass and high-pass filters, respectively. The two factors that should be considered for using the wavelet transform are the mother wavelet and the level of decomposition. Daubechies mother wavelet 'db7' has been suggested as a suitable candidate for wavelet analysis of PD pulses [18]. To determine the optimum level of decomposition, reference [4] proposes the maximum Shannon entropy using the energy of the approximation and wavelet coefficients. The maximum entropy defines the optimal decomposition level. Figure 2 shows the median of Shannon entropies for each decomposition level calculated using the PD pulses simulated by the AMTL model. It is observed that decomposition level 9 has the highest values overall. Thereby, it is chosen as the decomposition level in this work.

3.2. Feature Extraction

The accuracy of the classification model is directly influenced by the type of extracted features. Hence, precise feature analysis is required prior to classification. Traditionally, these features would be specified manually based on user experience and expertise. Such approaches can be time-consuming, and no one can guarantee that the selected features are the best types of features. However, an automated

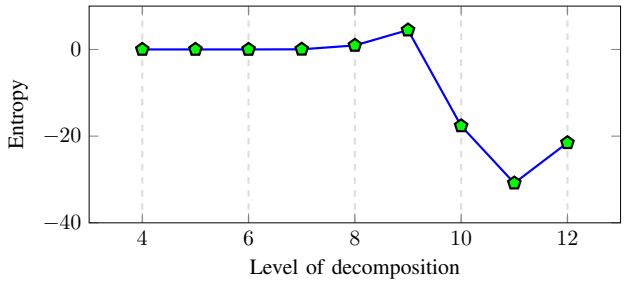

Figure 2. The median of Shannon entropies for each decomposition level calculated using the PD pulses simulated by the AMTL model.

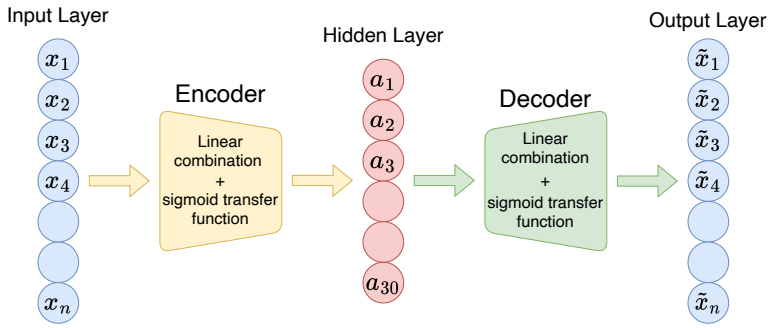

Figure 3. Autoencoder architecture with *n* input/output nodes and 30 hidden nodes.

alternative method can be feature extraction using an SAE [19]. In an SAE, the number of input and output nodes are equal, and the model should learn to optimize their similarity. The number of nodes in the hidden layer represents the number of features (see Figure 3) [19]. The hidden layer is calculated by passing the linear combination of the input layer z through a sigmoid transfer function defined as

$$h(z) = \frac{1}{1+e^{-z}}. \tag{2}$$

The output layer is constructed using a similar procedure.

The performance of this approach is compared with that derived from manual feature extraction later in this study. For this manual feature extraction, the features selected as the input of the classification model are categorized into three groups [10-13]:

- The energy of wavelet coefficients,
- Statistical parameters of wavelet coefficients, and
- Permutation entropy of wavelet coefficients,

where the wavelet coefficients that have been used include the last level of detailed coefficients and all levels of approximation coefficients. The energy of the j^{th} approximation coefficients is defined as:

$$E_j = \frac{\sum App_j^2}{numel(App_j)} \tag{3}$$

where App_j and $numel(App_j)$ are the j^{th} approximation coefficient vector and its number of elements, respectively. The statistical parameters employed in this work include the mean, standard deviation, skewness, and kurtosis [13]. The permutation entropies were calculated based on [12].

The first step was to extract features from the wavelet transform coefficients of the current pulses. The wavelet transform maps each signal into a vector consisting of its 9 approximation coefficients arrays and the first detailed coefficients array as the input to the SAE model. The optimum number of features would be the optimum size of the SAE hidden layer. In order to find this parameter, the model was launched on the input features for a different number of nodes of the SAE hidden layer, and the mean squared normalized error was measured for each trained model. The mean squared error with L2 and sparsity regularizers (msesparse) [20] was used as the loss function, and the encoder and decoder

transfer functions were sigmoid function . It is shown that there is a local minimum at 30 nodes. Furthermore, this size is not computationally expensive for the following analysis. At this point, msesparse reaches 0.0185 after 870 iterations.

4. RESULTS

Once the features were extracted, 70% of the data was assigned to the training set and 30% to the test set. A logistic regression, one-vs-all classifier was launched with a regularization of 0.001 and a maximum number of iterations of 250, as no considerable change could be seen after this iteration. The model could recognize the location of PDs with a training accuracy of 99.1% and a test accuracy of 97.9% after 250 iterations. The confusion matrix and the loss function are shown in Figure 4. To investigate the influence of the hyperparameters, the algorithm was repeated using a different number of regularizations, wavelet decomposition levels, and the number of hidden nodes in the SAE. Table 1 demonstrates the results for all the cases. No improvement in test accuracy can be seen in any of these modified models, for example when a value other than 0.001 was used for regularization. Regarding the number of features, changing the size of the hidden layer of the SAE by ±1 didn't change the training accuracy, but caused the test accuracy to drop. Although the performance of this method was considerably high, as another method, the features were extracted manually and used as input features for the classification model. The same classification model was employed with the group of featured proposed in Section 3.2 in order to see if in any of these scenarios, the classification performance improves. As shown in Table 2, the performance is reduced significantly in all the cases. The performances of other possible combinations of features, such as energy and entropy, were even worse in terms of test accuracy.

5. CONCLUSIONS

A transmission-line-based model was used to simulate an eight-disk transformer winding. Gaussian waveforms of variable rise time were injected into different locations of the winding to simulate partial discharge (PD) that occurs at that location. One end of the winding was open-circuit and the other end was grounded, where the current induced in the winding was recorded. The wavelet transform coefficients of the recorded currents were used as input to the proposed autoencoder/classifier model. The SAE extracted a deeper representation of the input data by mapping them into a latent space. The SAE model was employed with different hidden nodes to determine the optimal number of features. A logistic regression, one-vs-all classifier, was then employed to localize PD in the winding. The classification results showed an accuracy of 99.1% for the training set and 97.9% for the test set. The presented method was repeated using different values of hyperparameters, but no improvement was seen in the accuracy of the model. The classification was conducted without the SAE as well, where the features recommended in previous literature were used as input to the classifier. The accuracy of training and testing were compared with the previous case. The comparison showed a significant reduction in the performance of the classifier, which indicates the performance enhancement of the automated feature extraction over handcrafted feature extraction.

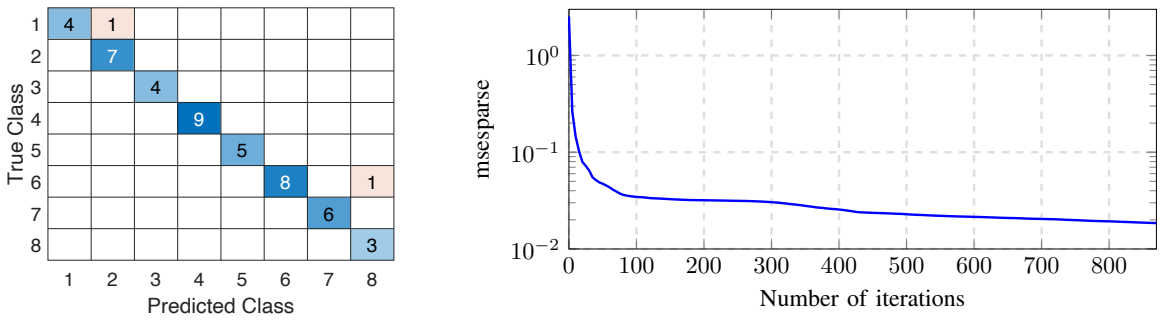

Figure 4. (a) The confusion matrix for the proposed model test set, (b) mean squared error with L2 and sparsity regularizers (msesparse) vs the number of iterations.

Table 1: The accuracy of the classifier with different value of hyperparameters.

	Training Accuracy (%)	**Test Accuracy (%)**
The proposed model	99.1	97.9
Level of decomposition = 8	99.1	93.8
Regularization = 0	100	93.8
Regularization = 0.0001	95.5	93.8
Number of Features = 29	97.3	91.7
Number of Features = 31	98.2	93.8

Table 2: The accuracy of the classifier with different value of hyperparameters with hand-crafted feature extraction.

	Training Accuracy (%)	**Test Accuracy (%)**
Statistical features only	86.8	78.3
Energy and statistical features	88.7	73.9
Entropy and statistical features	97.2	69.6
All features together	97.2	71.7

BIBLIOGRAPHY

[1] A. P. Marques, C. de Jesus Ribeiro, C. H. Azevedo, J. Lopes Dos Santos, F. de Carvalho Sousa, and L. da Cunha Brito, "Power transformer disruptions—A case study," *IEEE Electrical Insulation Magazine*, vol. 30, no. 2, Mar.-Apr. 2014.

[2] M. Mondal and G. B. Kumbhar, "Partial discharge localization in a power transformer: methods, trends, and future research," *IETE Technical Review (Institution of Electronics and Telecommunication Engineers, India)*, vol. 34, no. 5. Taylor and Francis Ltd., Sep. 03, 2017.

[3] J. M. Abdallah, "Power transformer windings partial discharge localization by transfer function," *International Journal of Electrical and Electronic Engineering* vol. 4. no. 6, 2010.

[4] D. Guillen, G. Idarraga-Ospina, E. Mombello, and S. Cabral, "Partial discharges location in transformer winding using wavelets and Kullback-Leibler divergence," *Electric Power Systems Research*, vol. 136, Jul. 2016.

[5] M. Nafar, T. Niknam, and A. Gheisari, "Using correlation coefficients for locating partial discharge in power transformer," *International Journal of Electrical Power and Energy Systems*, vol. 33, no. 3, Mar. 2011.

[6] K.-H. Fanchiang, Y.-C. Huang, C.-C. Kuo, "Power electric transformer fault diagnosis based on infrared thermal images using Wasserstein generative adversarial networks and deep learning classifier," *MDPI Electronics*, vol. 10, 2021.

[7] A. M. Gonçalves Júnior, H. de Paula, and W. do Couto Boaventura, "Practical partial discharge pulse generation and location within transformer windings using regression models adjusted with simulated signals," *Electric Power Systems Research*, vol. 157, Apr. 2018.

[8] S. M. H. Hosseini and P. R. Baravati, "Partial discharge localization based on detailed models of transformer and wavelet transform techniques," *Journal of Electrical Engineering and Technology*, vol. 10, no. 3, May. 2015.

[9] M. Gunawardana, F. Fattal, and B. Kordi, "Very fast transient analysis of transformer winding using axial multiconductor transmission line theory and finite element method," *IEEE Transactions on Power Delivery*, vol. 34, no. 5, Oct. 2019.

[10] M. S. A. Rahman, P. L. Lewin, and P. Rapisarda, “Autonomous localization of partial discharge sources within large transformer windings,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 23, no. 2, Apr. 2016.
[11] S. Wang, Y. He, B. Yin, W. Zeng, Y. Deng, and Z. Hu, “A partial discharge localization method in transformers based on linear conversion and density peak clustering,” *IEEE Access*, vol. 9, 2021.
[12] S. Govindarajan, M. Natarajan, J. A. Ardila-Rey, and S. Venkatraman, “Partial discharge location identification using permutation entropy based instantaneous energy features,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, 2021.
[13] A. M. G. Júnior, H. de Paula, W. C. Boaventura, S. M. A. Lopes, R. A. Flauzino, and R. A. C. Altafim, “Localisation of inter-layer partial discharges in transformer windings by logistic regression and different features extracted from current signals,” *IET Science, Measurement and Technology*, vol. 14, no. 10, Dec. 2020.
[14] J. A. Martinez-Velasco (editor), Power system transients: parameter determination. CRC press, 2017.
[15] A. Loubani, N. Harid, H. Griffiths, and B. Barkat, “Simulation of partial discharge induced EM waves using FDTD method—A parametric study,” *MDPI Energies*, vol. 12, no. 17, 2019.
[16] V. Javor and P. D. Rancic, “A channel-base current function for lightning return-stroke modeling,” *IEEE Trans Electromagnetic Compatibility*, vol. 53, no. 1, Feb. 2011.
[17] B. Kordi, J. LoVetri, and G. E. Bridges, “Finite-difference analysis of dispersive transmission lines within a circuit simulator,” *IEEE Transactions on Power Delivery*, vol. 21, no. 1, Jan. 2006.
[18] X. Ma, C. Zhou, and I. J. Kemp, “Interpretation of wavelet analysis and its application in partial discharge detection,” *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 9, no. 3, Jun. 2002.
[19] L. Duan, J. Hu, G. Zhao, K. Chen, J. He, and S. X. Wang, “Identification of partial discharge defects based on deep learning method,” *IEEE Transactions on Power Delivery*, vol. 34, no. 4, Aug. 2019.
[20] F. Khayatian, L. Sarto, and G. Dall’O, “Building energy retrofit index for policy making and decision support at regional and national scales,” *Applied Energy*, vol. 206, Nov. 2017.