International Journal Of
Industrial Electronics Control and Optimization
Print ISSN: 3517-2645
Online ISSN: 3568-2645
Homepage: https://ieco.usb.ac.ir/

Artificial Intelligence for Assessing Composite Insulator Pollution Level: A Study on Partial Discharge Characteristics

Hamid Reza Sezavar | Saeed Hasanzadeh

Department of Electrical and Computer Engineering, Qom University of Technology, Qom, Iran.[1,2,3]
Corresponding author's email: hasanzadeh@qut.ac.ir

Article Info

Article type:
Research Article

Article history:
Received: 31-March-2025
Received in revised form: 11-May-2025
Accepted: 16-May-2025
Published online: 21-March-2026

Keywords:
Composite insulator,
Gradient boosting machines,
Leakage current,
Partial discharge,
Pollution flashover,
Wavelet transformation.

ABSTRACT

Insulator pollution levels are critical for ensuring the operational stability and safety of power transmission systems. Traditional methods for detecting pollution are often invasive, inaccurate, and time-consuming. To address these issues, this study investigates the application of Artificial Intelligence (AI), specifically Gradient Boosting Machines (GBM), to classify insulator pollution levels based on Partial Discharge (PD) characteristics. We utilize a combination of time-domain and frequency-domain features extracted from PD signals to train a predictive model. The results indicate that the proposed model achieves a high classification accuracy, averaging between 92% and 95% across various contamination levels. Furthermore, the study analyzes the model's sensitivity to environmental factors, including humidity and Hydrophobicity Class (HC), revealing important insights that could influence classification performance. By employing this AI-driven approach, we aim to significantly enhance the efficiency of power grid maintenance, ultimately contributing to the long-term stability and reliability of transmission systems. The findings from this research underscore the potential of AI in revolutionizing pollution assessment methods and optimizing maintenance practices in power infrastructure.

I. Introduction

Pollution is one of the factors affecting the failure of insulators in the power system. Insulators in power system equipment lose their insulating properties as they are exposed to pollution. The accumulation of pollution, along with the addition of moisture, causes the formation of a conductive layer on the insulator and causes leakage current (LC) to flow on its surface. These signals gradually cause the appearance of small discharges on the surface [1]. These discharges are not detectable at first and cause the insulator's ageing, the insulating surface's loss, and premature failure. Partial discharges (PDs) generally occur on the surface of all insulators. In polymer insulators, more attention has been paid to dry band discharges and flashover discharges, and less attention has been paid to the analysis of PDs.

The performance of insulators is susceptible to the level of contamination. For this reason, analyzing and examining the level of contamination can help prevent flashovers. Many studies have been conducted to introduce methods for analyzing the level of contamination on insulators. The LC value is an important signal for determining the contamination level. In [2], it is shown that with increasing contamination levels, the LC value of the insulator increases, and the insulation performance is impaired. However, the LC value may not increase much at low humidity levels. As a result, the system has difficulty predicting the level of contamination. As a result, the idea of frequency analysis of LC was expanded instead of analyzing its value. This analysis began with the Fast Fourier Transform (FFT) and was expanded with the Wavelet Transform (WT). Initially, it was shown that the FFT analysis was also effective in low humidity [3]. In [4], the odd harmonics of LC were considered an example of frequency analysis, and a simple technique without the need for complex calculations was presented. In [5], frequency analysis was performed in dry conditions of the insulator with a humidity of 5%, and the ratio of the fundamental to the fifth and third harmonics was

investigated. In this case, it was shown that the ratio of the fundamental to the fifth harmonic is very effective in dry conditions and to the third harmonic in wet conditions. In [6], soluble and insoluble contaminations are investigated using FFT analysis. The main idea was that the voltage distortion increases proportionally to the contamination level due to the dry band arc. Therefore, this distortion is considered as an indicator of the contamination level. In [7], it is shown that the third and fifth harmonics of the insulation voltage play a fundamental role in determining the contamination level. In [8], the' energy value of individual harmonics is discussed to analyses and investigate the contamination level of composite insulators.

With the development of frequency analysis studies, models gradually moved towards using WT analysis. Both low and high frequencies were investigated over time in these investigations. In [9], discrete WT on insulators was started to estimate the contamination level, and in [10], frequency analysis was extended to index the contamination level. In [11], the WT was investigated in insulators with salt fog conditions, and in [12], this topic was expanded to ceramic insulator. Along with WT, Artificial Intelligence (AI) has also been significantly developed in estimating pollution levels in recent studies. Recent advancements in AI provide opportunities to improve the accuracy and efficiency of insulator pollution level assessments. For example, Artificial Neural Networks (ANN) have been used in the analysis and investigation of insulator pollution conditions [13, 14], Support Vector Machines (SVM) [15, 16] in the assessment of pollution levels, and other methods related to AI to estimate pollution level density [17, 18] and fault diagnostic in transformers [19] have been significantly developed.

However, the most important gap in studies of pollution level determination is the lack of use of physics-based analyses of the problem. The physics of pollution on the insulator surface causes PD on the surface [20, 21], and therefore, this idea can be a starting point for determining the pollution level. In [22], it is mentioned that the intensity and power of PD can depend on the pollution level. Also, in [23, 24], it is shown that PD signals have significant changes in high frequencies with changing weather conditions. PDs, mainly in insulators, are the initiators of flashover and are often invisible and can only be calculated from the LC signal. Therefore, a suitable relationship between the physics at the contamination level and the initiation of PDs can be obtained.

This study explores the application of Gradient Boosting Machines (GBM), a powerful machine learning technique, to classify insulator pollution levels based on PD characteristics. These characteristics are extracted from both the time-domain and frequency-domain features of PD signals. The primary objective is to develop an AI model capable of assessing insulator pollution levels in a fast, non-invasive, and reliably. This AI model aims to bridge the gap between physics-based analysis and machine learning approaches by integrating the underlying physical principles governing PD and pollution effects on insulators. It has been shown that PDs have features in the range of 5 to 15 kHz that relate to pollution levels. By leveraging the characteristics of PD signals, the GBM will classify pollution levels more accurately, accounting for variations in environmental conditions, such as humidity and contamination types. Integrating GBM with physical insights into PD behavior significantly advances the assessment of insulator pollution levels. Furthermore, this study will incorporate a comprehensive dataset that includes various insulator types and environmental scenarios to ensure robust model training and validation. The findings are expected to provide a novel framework for real-time monitoring of insulator health, enhancing predictive maintenance strategies in power systems, and ultimately improving their reliability and performance.

II. Laboratory setup

The experiments were conducted in a cubic climate chamber measuring 2 meters. The power supply included a step-up transformer rated at 100 kV and a protective circuit. A capacitive voltage divider with a voltage ratio 10,000:1 was used to measure the applied voltage. The LC was also recorded through the voltage across a shunt resistor connected in series with the insulators. PD signals were gathered using LC peak detection and filtered to ensure accurate readings. Figure 1 illustrates the schematic of the testing circuit and the climate chamber.

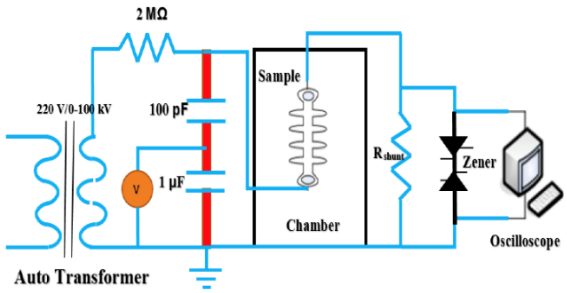


Fig. 1. Schematic view of the test setup.

TABLE I PARAMETERS OF THE SAMPLE TEST

Arcing Distance (cm).	Diameter of sheds (cm)	Num. of sheds	Creepage distance (cm)
33	10	6	85

All tests utilized 20 kV composite insulators from the same silicone rubber (SiR) material. The two types of pollution evaluated were equivalent salt deposit density (ESDD) and non-soluble deposit density (NSDD). The NSDD consisted of standard kaolin material. The ranges for ESDD and NSDD examined in this study were approximately 0 mg/cm² to 0.5 mg/cm² and from 0 mm to 5 mm in thickness, respectively, covering all levels of pollution classified as low, medium, high, and very high according to IEC 60815. The samples were contaminated by immersing them in a pre-prepared solution of distilled water mixed with kaolin to simulate NSDD. An average pollution ratio from top to bottom was calculated to be 0.96, indicating a relatively uniform distribution of the pollution layer. The specifications for the test insulator construction are detailed in Table 1 [25].

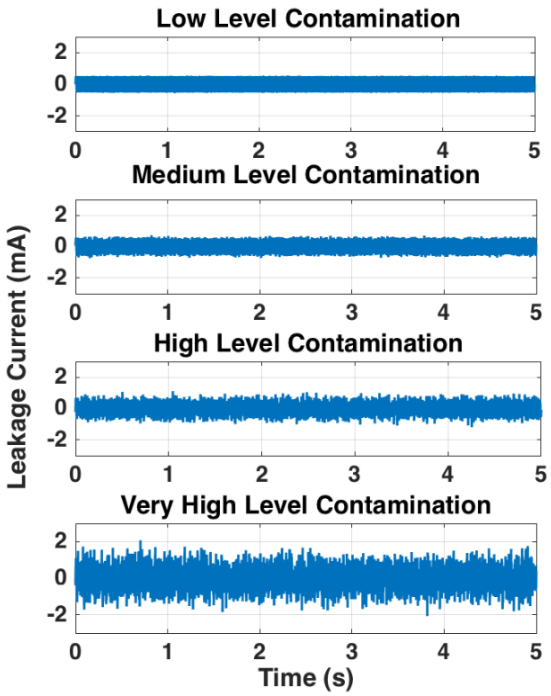
The input voltage was incrementally increased at a 1 kV/min rate throughout the testing phase, following IEC 60243-1, until dry-band arcing was noted. This protocol is used to achieve the PD signals because the PD signal typically appears before dry band arcing. During the experimentation, contaminants were sprayed into the chamber, maintaining a constant equivalent salt deposit density (ESDD) using a solution made of distilled water and sodium chloride (NaCl). During the test, PD activities were continuously monitored, and the signals were recorded to assess the insulation performance and detect any insulation degradation.

III. Methodology

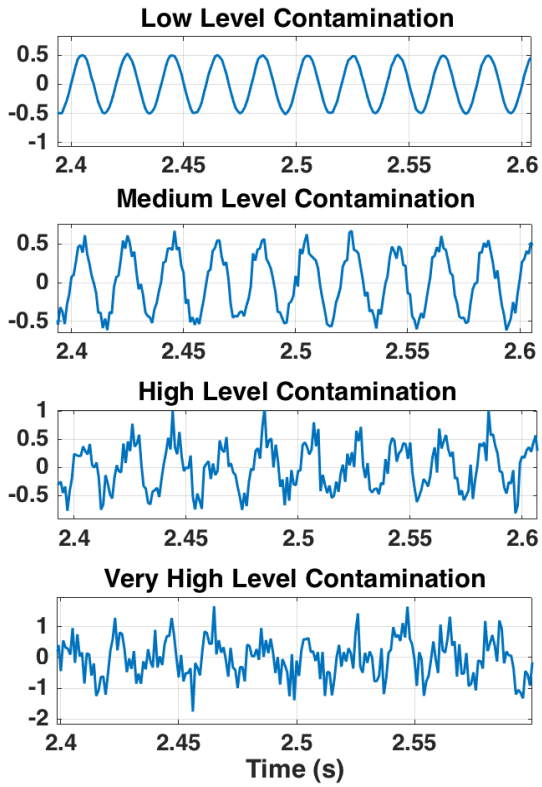
A. PD Characteristics

The results in Figure 2 represent the four contamination levels according to the conditions described in section 2. The results are obtained based on low humidity and medium-class hydrophobicity. As Figure 2(b) shows, with increasing contamination levels, insulator LC, including peak signals, appears on the surface. Despite the difficulty of detecting PD, considering the changes in insulator LC with increasing contamination level in the insulator, signs of PD detection can be found. These changes can be a starting point for PD analysis and evaluating the insulator contamination level.

For a more detailed analysis, Figure 3a shows the relationship between the LC signal and its frequency analysis. The results of the Fourier transform analysis show that the frequency amplitude of the PD signal increases with the increase in the contamination level. However, a significant difference is observed in the 5 to 15 kHz range in figure 3b. With increasing contamination levels, the amplitude of the Fourier tidal signal in this range increases. This point is very valuable because it can be a relationship to evaluate the insulator contamination level based on the changes in the PD signal on the insulator surface. Further, based on the analysis of the frequency-time behavior in the stated range, features can be obtained to correctly assess the pollution level.

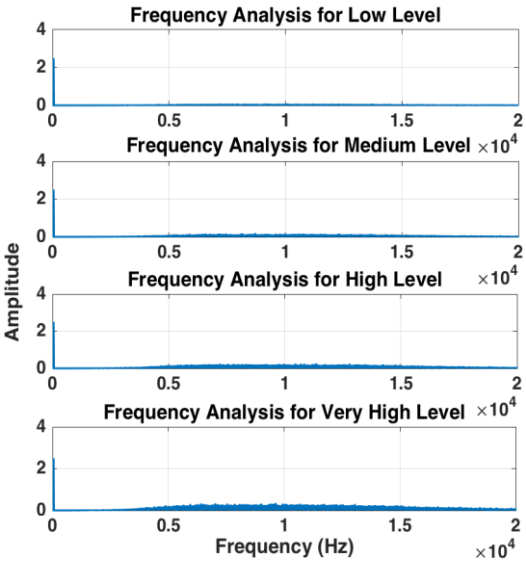


(a)

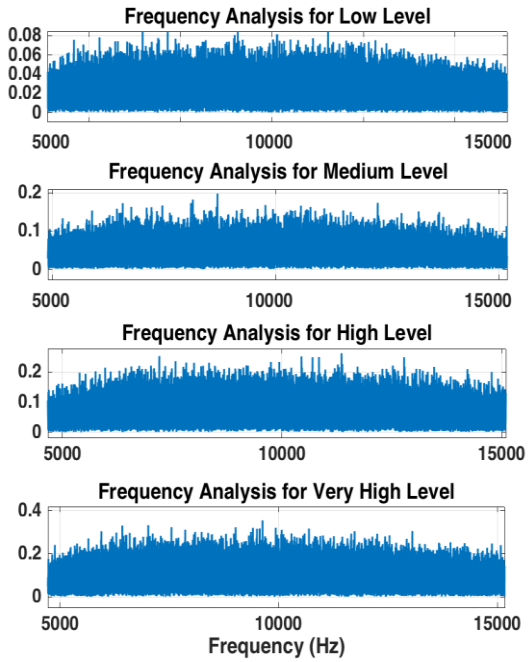


(b)

Fig. 2. (a) LC signals and (b) PD appearing, based on the four contamination levels.



(a)



(b)

Fig. 3. a) Frequency analysis of LC signal. b) Amplitude of frequency analysis in 5000 to 15000 Hz

B. Feature Extractions

The current study employed Wavelet Packet Transform (WPT) for feature extraction, which serves as a broader representation of the wavelet transform. In each iteration, WPT executes a new decomposition based on coefficients obtained from its prior iterations. This implies that the final number of coefficients is contingent upon the number of decompositions conducted [26]. When applying a wavelet decomposition at the wavelet packet node level (WP), the segmentation of approximation coefficients results in a tree structure composed of two vectors: one representing the approximation coefficient vector and the other designated as the detailed vector. Information lost during the approximation process is encapsulated in the previously mentioned coefficients, forming a new vector. In this context, successive details are not reanalyzed [27].

The WP function can be expressed as:

$$W_{j,k}^{n}(t) = 2^{\frac{j}{2}}(2^{j}t - k) \tag{1}$$

where j is a scaling parameter, k denotes the translation operator, and n represents the oscillation parameter. The initial two WP functions for n=0, 1 are defined as follows:

$$W_{0,0}^{0}(t) = \emptyset(t) \tag{2}$$

$$W_{0,0}^{1}(t) = \psi(t) \tag{3}$$

The first function in Equation (2) is the scale function, while the second is the wavelet function [28]. Subsequent functions for n = 2, 3,…,N can be specified using these relationships:

$$W_{0,0}^{2n}(t) = \sqrt{2\sum_{k}\delta(k)W_{1,k}^{n}(2t-k)} \tag{4}$$

$$W_{0,0}^{2n+1}(t) = \sqrt{2\sum_{k}\xi(k)W_{1,k}^{n}(2t-k)} \tag{5}$$

where $\delta(k)$ refers to a low-pass filter and $\xi(k)$ denotes a high-pass filter, both associated with a predefined scaling function and the mother wavelet function. The coefficients $\Omega_j^n(t)$ can be derived by considering the product of the functions $x(t)$ and $W_{j,k}^{n}(t)$:

$$\Omega_j^n(t) = \int_{-\infty}^{+\infty} x(t)\, W_{j,k}^{n}(t)dt \tag{6}$$

Each coefficient WP corresponds to a specific frequency level. While the wavelet transform focuses on low-frequency components, WPT encompasses all elements. This results in low and high-frequency components, known as low and high approximations. In applying WPT, factors such as entropy, energy, and variation must be accounted for during the WP calculation process. Energy serves to characterize distinct classes and, in the proposed methodology, encapsulates failure indicators related to the condition of the insulator. Fluctuations in energy correspond to specific failure types similar to those outlined in [29]. The signal undergoes decomposition into j levels, which establishes orthogonal subspaces from where frequency components can be calculated using the following:

$$\mathrm{E}_j^n = \sum_{k}\left|\Omega_j^n(k)\right|^2 \tag{7}$$

For energy normalization within each frequency band, the distribution percentage for the energy component is expressed as:

$$\mathrm{e}_j^n = \frac{\mathrm{E}_j^n}{\sum_{n=1}^{2j}\mathrm{E}_j^n} \tag{8}$$

The relative energy of the vector illustrates temporal development considering low and high-frequency subspaces. Changes in this distribution pattern reflect energy flow and can reveal identifiable patterns. A binary optimal

value can be determined by the previously discussed tree structure formed by the segmentation of approximation coefficients. This creates new subdivisions (sub-trees) based on the entropy criterion. Depending on the application, these resulting sub-trees may be significantly smaller than the original. This method seeks to identify a minimum criterion to facilitate an efficient algorithm [28].

Figure 4 shows the energy results at four pollution levels at j=3, 4, 5, and 6 based on the presented wavelet model. According to the stated points, these four levels correspond to the energy received from the signal in the frequency range of 5 to 15 kHz. This interval is the same interval that was determined in the frequency analysis and can be effective in evaluating different pollution levels. However, considering the closeness of the results, a more appropriate model can be proposed. Entropy analysis is used considering the PD signal's random nature.

Considering the variability of PD based on the pollution level, the frequency-time analysis can be performed in the 5 to 15 kHz range. This analysis can be performed based on the wavelet transform model. According to [29], the wavelet packet is divided into low and high-frequency bands. In order to achieve a frequency of 5 to 15 kHz, the signal decomposition should be at levels 3, 4, 5, and 6, Considering that the LC signal acts sinusoidal, the mother function can be used based on the three-layer decomposition. In this situation, the uncertainty framework is adopted for probabilistic treatment and is defined as a logarithmic function $H(p_1, \dots, p_n)$, given by:

$$H(p_1, \dots, p_n) = -\sum_{i=1}^{n} p_i log p_i \quad (9)$$

where $p_i$denotes the probability of occurrence in an event i, the entropy reflects the probabilistic uncertainty in a probability distribution [30]. Figure 5 shows the entropy results based on the change in the pollution level. The results of the figure clearly show the change in the entropy level based on the change in the pollution level. This indicates the appropriate performance of the model in the analysis of PD based on the pollution level.

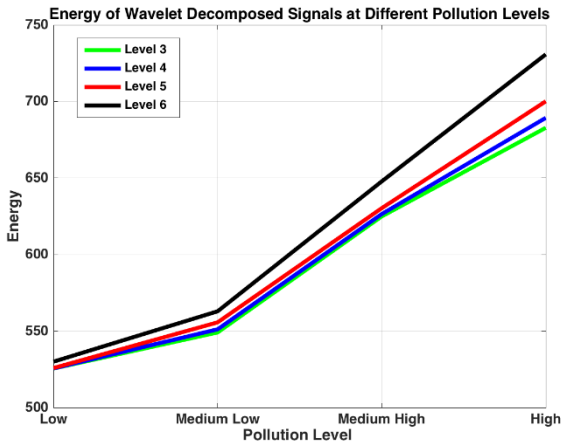


Fig. 4. Energy results at four pollution levels

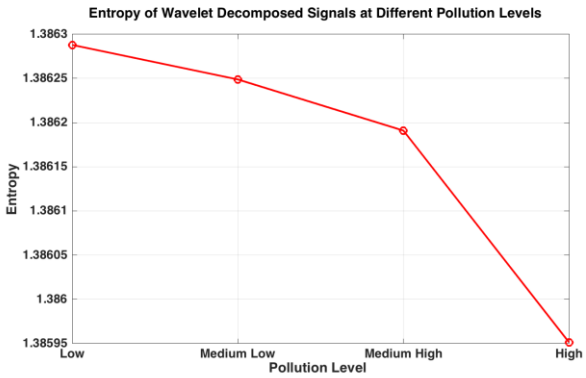


Fig. 5. Entropy results based on the change in the pollution level.

C. *Gradient Boosting Machine Model*

GBM is an advanced ensemble learning method that constructs a series of weak learners, typically decision trees, to create a strong predictive model. Rather than building a single complex model, GBM works by sequentially training multiple models, where each new model focuses on correcting the errors made by the previous ones. In each iteration, the model adjusts the weights assigned to the training samples; misclassified samples gain higher weights, while correctly classified samples are given lower weights. This iterative adjustment continues until a specified number of models have been trained or until the model's error minimizes to an acceptable level [31].

For this study, XGBoost, which stands for Extreme Gradient Boosting, is implemented. XGBoost is a highly optimized and efficient version of the traditional GBM, recognized for its exceptional speed and accuracy in making predictions. It includes numerous enhancements, such as parallel processing, tree pruning, and regularization techniques, contributing to its superior performance over standard GBM implementations. The model was trained using the extracted features derived from PD signals, where each feature vector encompasses specific signal characteristics, allowing the model to learn patterns indicative of varying pollution levels.

To ensure the best results, hyperparameters for XGBoost were meticulously tuned, focusing on critical parameters such as the learning rate, which controls the step size at each iteration; maximum depth, which dictates the complexity of the trees; and the number of estimators, which determines how many trees to build. This tuning process was conducted using cross-validation techniques, allowing us to evaluate the model's performance on unseen data and mitigate the risk of overfitting. By optimizing these hyperparameters, it is aimed to achieve a balance between model complexity and generalization, ultimately enhancing the model's predictive capability for assessing pollution levels in composite insulators through PD signal analysis. The neural network model employed in this study is a feedforward neural

network with five hidden layers containing 1000 neurons. The hidden layer uses ReLU activation, and the output layer employs a softmax activation for classification. The network was trained using a learning rate of 0.001 and the Adam optimiser, with early stopping to prevent overfitting. In comparison, the XGBoost model was implemented with hyperparameters optimised through grid search and cross-validation. The selected hyperparameters are: 100 estimators, maximum tree depth of 6, learning rate of 0.05, subsample ratio of 0.8, column sample by tree of 0.8, gamma of 0.1, and minimum child weight of 4. In gradient boosting, the key idea is to minimize the loss function *L(y,F(x))* using weak learners, each represented by a function $F(x)$. The updated model at iteration m can be expressed as: [31].

$$F_m(x) = F_{m-1}(x) + \nu h_m(x) \quad (10)$$

where $F_m(x)$ is the prediction from the model at the m_{th} iteration. $F_{m-1}(x)$ is the prediction from the previous iteration. ν is the learning rate, and $h_m(x)$ is the new a decision tree. The objective of gradient boosting is to minimize the loss function, which could be the mean squared error for regression tasks or log loss for binary classification tasks. For a regression problem, the loss function can be defined as:[31]

$$L(y, \hat{y}) = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \quad (11)$$

which y_i is the actual target value. $\hat{y}_i$ is the predicted value from the model.

The final prediction of the XGBoost model can be calculated as:

$$\hat{y} = F(x) = \sum_{m=1}^{M} h_m(x) \quad (12)$$

where M is the total number of trees. In the context of decision trees (which are the weak learners in GBM and XGBoost), entropy is used to measure the impurity of a node in the tree. The idea is to split the data based on features such that the entropy is minimized, leading to more homogeneous subsets. In XGBoost, while the primary objective function often utilizes squared loss for regression or log loss for classification, it can be uses entropy for classification tasks. The loss function can be expressed in terms of log loss and simulated using cross-entropy. For classification, the cross-entropy loss $L(y, \hat{y})$ can be defined as:

$$L(y, \hat{y}) = -[y log(\hat{y}) + (1 - y)\log(1 - \hat{y})] \quad (13)$$

y is the true label (0 or 1) and $\hat{y}$ is the predicted probability. When combining the proposed model with the previous concepts, calculating gradients for XGBoost will involve the derivatives of the cross-entropy loss, effectively using entropy principles. The flowchart in Figure 6 is a representation of the proposed model.

Given the utilized XGBoost, it has leveraged the inherent feature importance scores provided by the algorithm. XGBoost computes feature importance based on how frequently a feature contributes to reducing the loss function across all trees (Tree-based Feature Importance). It selects features with high-importance scores as candidates for the optimal input set. Cross-validation techniques have been utilized to ensure the robustness of the feature selection. Training the model on different subsets of the data and validating it on complementary subsets has verified that the selected features consistently provided good predictive performance.

IV. Results and Discussion

In order to analyze the presented model, according to the flowchart of Figure 6, 400 test data samples from various contamination levels have been examined. One hundred data samples were examined at each contamination level. While The training dataset is derived from multiple tests, covering pollution levels ranging from 0 mg/cm² to 0.5 mg/cm² for ESDD and from 0 mm to 5 mm in thickness for NSDD, simulating low, medium, high, and very high pollution states as classified by IEC 60815. The signal data collected included Partial Discharge (PD) characteristics, which are processed to extract features. The testing dataset comprised additional measurements taken under the same experimental conditions but utilized different insulator samples to ensure the generalizability of the proposed model. Testing was conducted after training to evaluate the model's performance accurately. The training and testing data are 70% and 30% of all data, respectively.

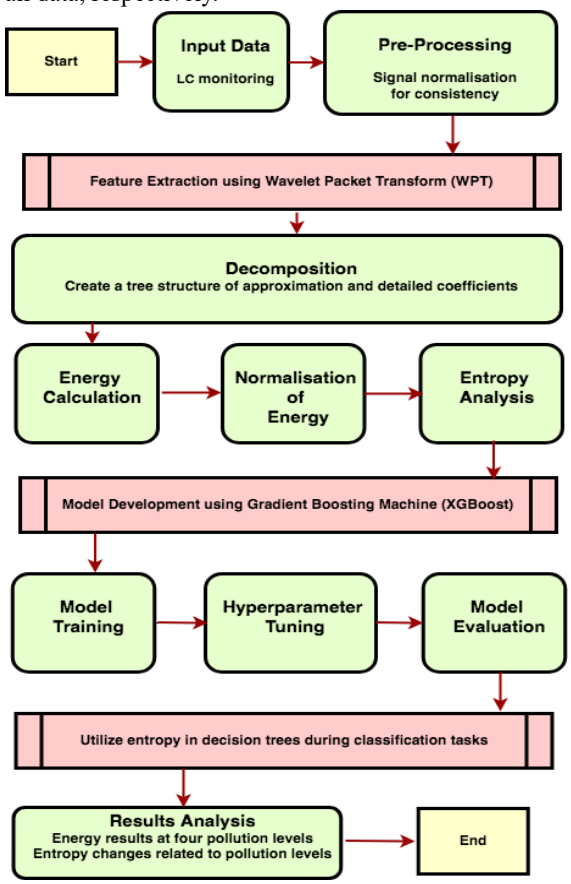


Fig. 6. Flowchart of the presented method

Figure 7 shows the trend of the loss function in the proposed method. The heat map matrix has been drawn as shown in Figure 8. The results of the figure show that the results match 93%, 95%, 92%, and 94% on average for low, medium, high, and very high contamination levels, respectively. The slight deviations from 95% accuracy can be attributed to inherent variability within the test data samples. Each contamination level may exhibit variations in characteristics that are less pronounced, making it challenging for the model to achieve uniform accuracy across all levels. The nature of insulator pollution can result in complex interactions that affect the partial discharge characteristics, which in turn impact the model's classification capabilities. These complexities are particularly evident at higher contamination levels where nuanced differences in the PD signals may influence the model's predictions. The model is trained to generalize across various contamination scenarios. While it performed well (as indicated by the accuracies), the less-than-perfect scores suggest room for improvement. This is typical in real-world scenarios where a model may not fully capture all the variations in the data. In order to comprehensively analysis the model, the precision ratio (Pr) has been examined as a suitable data evaluation model. The accuracy ratio is between the number of samples evaluated at each contamination level and the total number of data. Also, the ratios in Figure 8 are considered the recall rate (Rr). In order to obtain accurate information, the F-score index has been used. The formula of this index is as follows:

$$F - Score = 2 \times \frac{P_r \times R_r}{P_r + R_r} \tag{14}$$

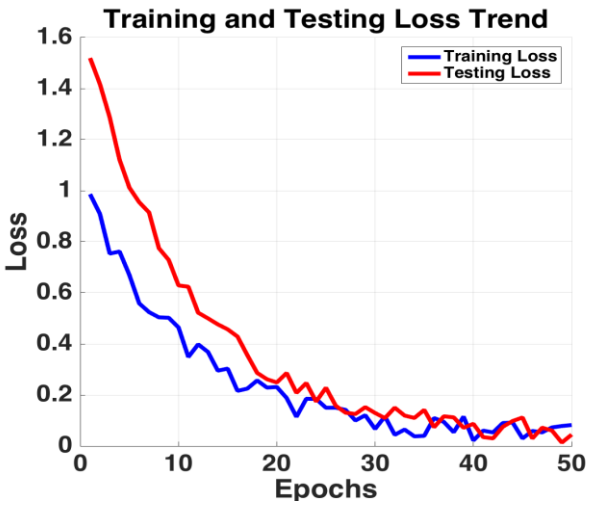


Fig. 7. The trend of the loss function in the proposed method in training and testing data

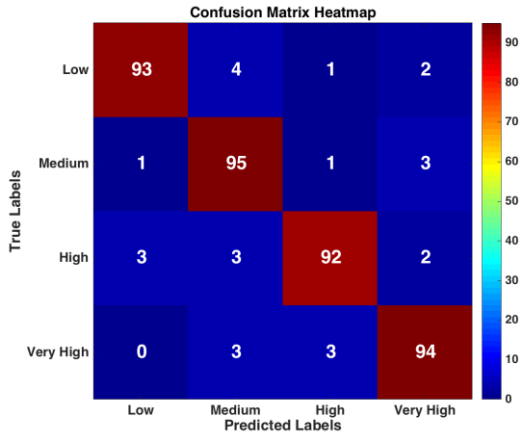


Fig. 8. The heat map matrix of the proposed method's results

A. Sensitivity analysis of pollution level

Table II presents the classification results of the predictive model used to assess pollution levels, revealing a strong overall performance characterized by high accuracy rates ranging from 92% to 95% across different pollution classifications (I to IV). The model demonstrates particularly robust performance for Pollution Level II with an accuracy of 95%. In comparison, high-level pollution (Pollution Level III) shows a slightly lower accuracy of 92%, indicating potential challenges in distinguishing this category. The average accuracy across all levels remains high, with values of 96%, 90%, 95%, and 93%, respectively, demonstrating the model's reliability. Furthermore, the F-scores, which range from 0.924 to 0.945 for the different pollution levels, emphasize a well-balanced approach in achieving both precision and recall rates, indicating that the model effectively minimizes false positives. Despite these commendable results, the assessment matrix reveals areas for improvement, particularly for high and very high-level (III and IV), where misclassifications occurred. Addressing these inaccuracies through model refinement and additional training data could enhance the model's effectiveness, ultimately leading to better classification of pollution levels and contributing to more informed environmental management strategies.

TABLE II ASSESSMENT RESULTS OF POLLUTION CLASS

Pollution Level		Assessment Levels				Total	Rr
		I	II	III	IV		
Actual	I	93	4	1	2	100	93%
	II	1	95	1	3	100	95%
	III	3	3	92	2	100	92%
	IV	0	3	3	94	100	94%
Total		97	105	97	101	-	-
Pr		96%	90%	95%	93%	-	-
F-Score		0.945	0.924	0.935	0.936	-	-

B. Sensitivity analysis of humidity

In order to verify the results, a humidity analysis was performed on the test insulator, revealing noteworthy insights into the model's sensitivity to varying humidity levels. As shown in Table III, the F-scores for pollution levels I to IV demonstrate a general trend of decreasing performance with increasing humidity; however, all F-score values remain above 90%, indicating robust model reliability. Specifically, for Pollution Level I, F-scores decreased from 94.5% at 60% humidity to 90.8% at 90% humidity, suggesting that higher humidity may complicate the classification of low pollution levels due to increased electrical discharges. Conversely, Pollution Level III exhibited a peak F-score of 95.3% at 70% humidity before declining at higher levels, indicating the model's adeptness at classifying moderate pollution under optimal conditions. The important point is the uniform trend in pollution levels III and IV in all humidity levels, which shows the model's reliability in high and very high pollution levels in different humidity. Despite the increased misclassifications at higher humidity, the model's ability to maintain strong performance metrics underscores its resilience.

C. Sensitivity analysis of HC level

In order to evaluate the model's performance concerning the age of the insulators, an HC analysis is conducted, as detailed in Table IV. The results demonstrate that the proposed model effectively detects older insulators, evidenced by consistently high F-scores across all pollution levels. Specifically, for Pollution Level I, the F-scores range from 90.1% for HC level 1 to 96.4% for HC levels 4 and 6, indicating that older insulators are more susceptible to partial discharges and, therefore, can be detected more readily by the model. In contrast, new insulators exhibit reduced F-scores, with the highest value at 95.7% for Pollution Level II at HC level 6, indicating their lower susceptibility to partial discharges and flashover events. While newer insulators tend to produce fewer misclassifications, which leads to lower F-scores, the model remains remarkably reliable when assessing older insulators. These findings underscore the model's potential for practical applications across varying contamination levels, conditions, and insulator ages.

TABLE III ASSESSMENT RESULTS OF HUMIDITY SENSITIVITY ANALYSIS

F-Score %		Humidity			
		60%	70%	80%	90%
Pollution Level	I	94.5	93.5	91.2	90.8
	II	92.4	92.7	91.1	90.2
	III	93.5	95.3	94.2	93.1
	IV	93.6	95.1	93.8	92.3

TABLE IV ASSESSMENT RESULTS OF HC SENSITIVITY ANALYSIS

F-Score %		HC			
		1	3	4	6
Pollution Level	I	90.1	94.5	96.4	96.4
	II	88.3	92.4	95.4	95.7
	III	89.2	93.5	94.3	94.8
	IV	85.4	93.6	93.2	94.4

D. Validation model

The comparative analysis of the proposed model, Support Vector Machine (SVM) [32], and Neural Network (NN) [33] reveals the strengths and weaknesses of each approach in classifying pollution levels based on synthetic data. The results indicate that the proposed model achieved the highest accuracy among the evaluated models, with performance reaching approximately 94.45%. The out-of-bag (OOB) error estimates provided during training further reinforce the model's reliability, indicating how well the model is likely to perform on unseen data.

In contrast, while generally regarded as a robust classification tool, the SVM model has a lower accuracy of around 91.13%. In addition, the use of the SVM method also has limitations. For example, there is complex, multi-dimensional data without carefully tuning hyperparameters and kernel functions. The ability of SVM to create decision boundaries is highly reliant on selecting these parameters, which may not always reflect the optimal separation for non-linear datasets. The Neural Network model, trained with a feedforward architecture containing ten hidden neurons, displayed comparable results to the SVM, achieving around 88.73% accuracy. However, its performance illustrates common challenges associated with neural networks, including the necessity for proper initialization, training time, and the risk of overfitting when working with synthetic data. While neural networks excel in capturing intricate relationships in larger datasets, their reliance on extensive tuning and the quality of training data can sometimes impede their efficiency in more straightforward tasks compared to ensemble methods.

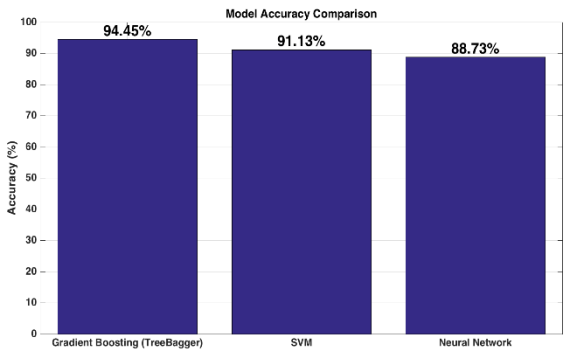


Fig. 9. The comparative analysis of the proposed model, Support Vector Machine, and Neural Network

The overall performance of the proposed model suggests its suitability for classification tasks in diverse contexts, including pollution monitoring, where adaptability to different conditions is vital. The model's resilience against overfitting and its ability to handle noisy data make it a powerful option for real-world applications. Given the nature of pollution data, which can be influenced by various environmental factors, a model that balances accuracy, interpretability, and computational efficiency is essential for effective management. While the Gradient Boosting Model stands out in this comparison, it is essential to recognize that no single model is universally superior. The choice of model ultimately depends on specific application requirements, computational resources, and the characteristics of the dataset being analyzed. Future work should explore enhancements to each model, such as fine-tuning hyperparameters for the SVM and NN or investigating more advanced ensemble methods to improve classification results even further.

The proposed model demonstrates robust performance in scenarios where multiple insulators operate in parallel, even in the presence of signal coupling and interactions. Advanced signal processing techniques have been integrated to enhance the model's capability for operating in coupled environments, allowing it to analyze mixed signals and identify individual partial discharge signatures effectively. Techniques such as time-frequency analysis aid in discerning the frequency components of PD signals over time, enabling the separation of signals from multiple insulators. Additionally, the model employs feature extraction methods to analyze key characteristics of PD signals—such as peak amplitudes, rise times, and energy levels—all of which support accurate pollution level classification even amidst coupling interference. Furthermore, using Gradient Boosting Machines (GBM) enhances the model's proficiency in managing complex datasets and interactions between input variables. This strategic design ensures that the proposed model maintains high classification accuracy and reliability across diverse operational conditions, effectively addressing the challenges of parallel insulator networks.

V. Conclusions

This study introduces a novel AI-driven approach for assessing insulator pollution levels using Gradient Boosting Machine (GBM) and Partial Discharge (PD) characteristics, demonstrating high accuracy across diverse pollution scenarios. The developed model offers a rapid, non-invasive, and reliable alternative to traditional inspection methods, which are often hindered by environmental conditions and physical contact limitations. Importantly, the computational efficiency and robustness of the model suggest strong potential for real-time, online condition monitoring in operational power systems. Implementing such an AI-based system could enable continuous health assessment of insulators, facilitating early detection of pollution-related issues and enabling proactive maintenance, ultimately reducing the risk of outages and enhancing system reliability. In order to prevent flashover in contaminated insulators, the proposed method, based on the detection of LC and PD characteristics, is capable of detecting the contamination level. In this case, the online monitoring system can predict a fault's occurrence before it occurs. Future advancements, including dataset expansion under various environmental conditions and the integration of the model into existing monitoring infrastructure, are essential steps toward realizing practical, real-time deployment in field conditions.

REFERENCES

[1] A. Ghaedi, R. Sedaghati, M. Mahmoudian, and S. Bazyari, “De-Noising of Partial Discharge Signals in HV XLPE Cables by Reference Noise based on the Wavelet Transform,” *International Journal of Industrial Electronics Control and Optimization,* vol. 6, no. 4, pp. 291-306, 2023, doi: 10.22111/ieco.2023.46346.1498

[2] T. Kim and J. Yi, “Application of hydrophobic coating to reduce leakage current through surface energy control of high voltage insulator,” *Applied Surface Science,* vol. 578, p. 151820, 2022, doi: 10.1016/j.apsusc.2021.151820

[3] S. N. Razavinia *et al.*, “Numerical Simulation of Electron Cyclotron Resonance Discharge Generated by Permanent Magnets,” *Power, Control, and Data Processing Systems,* vol. 2, no. 1, 2025, doi: 10.30511/pcdp.2025.718447

[4] M. E. Ibrahim and A. M. Abd-Elhady, “Rogowski coil transducer-based condition monitoring of high voltage insulators,” *IEEE Sensors Journal,* vol. 20, no. 22, pp. 13694-13703, 2020, doi: 10.1109/jsen.2020.3005223

[5] S. Deb, R. Ghosh, S. Dutta, S. Dalai, and B. Chatterjee, “Effect of humidity on leakage current of a contaminated 11 kV Porcelain Pin Insulator,” in *2017 6th International Conference on Computer Applications In Electrical Engineering-Recent Advances (CERA)*, 2017, pp. 215-219: IEEE, doi: 10.1109/CERA.2017.8343329

[6] L. Jin, Z. Xu and S. Zhang, “A pre-warning method of contamination flashover based on the leakage current of insulators in dry condition,” *2017 International Symposium on Electrical Insulating Materials (ISEIM)*, Toyohashi, Japan, 2017, pp. 757-760, doi: 10.23919/ISEIM.2017.8166599.

[7] U. Benisheikh, Z. Abdul-Malek, and A. Sahito, “Classification of leakage current odd harmonic component of aged insulator surface condition under different relative humidity using crest factor,” *Archives of Electrical Engineering,* pp. 1069-1085, 2024, doi:10.24425/aee.2024.152111

[8] L. Zhengfa *et al*., “Study on leakage current characteristics and influence factors of 110kV polluted composite insulators,” *2018 12th International Conference on the Properties and Applications of Dielectric Materials (ICPADM)*, Xi'an, China, 2018, pp. 896-900, doi: 10.1109/ICPADM.2018.8401173.

[9] M. A. Amini and A. R. Sedighi, “A new procedure for determination of insulators contamination in electrical distribution networks,” *International Journal of Electrical Power & Energy Systems,* vol. 61, pp. 380-385, 2014, doi: 10.1016/j.ijepes.2014.03.034.

[10] M. Amini, A. S. Anaraki and M. Sadeghi, "High frequency components of feeder current as diagnostic tool to study contamination conditions of outdoor insulators in power distribution networks," *2013 21st Iranian Conference on Electrical Engineering (ICEE)*, Mashhad, Iran, 2013, pp. 1-6, doi: 10.1109/IranianCEE.2013.6599890.

[11] S. Ahmad, R. Ahmed, R. Abd Rahman, A. Ullah, A. Jamal, and R. Akram, "A comprehensive study of nano/micro fillers on silicone rubber insulators: Electrical, mechanical, and thermal characterization," *Results in Engineering,* vol. 25, p. 103654, 2025, doi: 10.1016/j.rineng.2024.103654.

[12] B. Cao, Y. Liu, Z. Li, S. Shen, L. Wang and Z. Wang, "Assessment of Insulator Pollution Degree Based on Contamination Moisture With Temperature Change," in *IEEE Sensors Journal*, vol. 22, no. 21, pp. 21172-21178, 1 Nov.1, 2022, doi: 10.1109/JSEN.2022.3206215.

[13] Y.-T. Lin and C.-C. Kuo, "Real-time salt contamination monitoring system and method for transmission line insulator based on artificial intelligence," *Applied Sciences,* vol. 14, no. 4, p. 1506, 2024, doi: 10.3390/app14041506.

[14] S. Gao *et al*., "Prediction method of leakage current of insulators on the transmission line based on BP neural network," *2018 IEEE 2nd International Electrical and Energy Conference (CIEEC)*, Beijing, China, 2018, pp. 569-572, doi: 10.1109/CIEEC.2018.8745839.

[15] Z. Fei, Y. Li, and S. Yang, "Partial Discharge Pattern Recognition Based on an Ensembled Simple Convolutional Neural Network and a Quadratic Support Vector Machine," *Energies,* vol. 17, no. 11, p. 2443, 2024, doi: 10.3390/en17112443

[16] Q. Zhu *et al.*, "Research on partial discharge pattern recognition of transformer based on KPCA and JS-SVM algorithm," in *Journal of Physics: Conference Series*, 2025, vol. 2963, no. 1, p. 012013: IOP Publishing, doi: 10.1088/1742-6596/2963/1/012013

[17] H. Jung, Y.-T. Kim, S.-K. Lee, and J.-h. Ahn, "Study on Deep-Learning Model for Phase Resolved Partial Discharge Pattern Classification Based on Convolutional Neural Network Algorithm," *Journal of Electrical Engineering & Technology,* vol. 20, no. 1, pp. 873-878, 2025, doi: 10.1007/s42835-024-01967-9

[18] C. W. Priananda, H. A. Illias, W. J. K. Raymond and I. M. Y. Negara, "Hybrid Deep Learning Models for Enhanced Classification of Phase-Resolved Partial Discharge Patterns from High-Voltage Rotating Machine Insulation," in *IEEE Transactions on Dielectrics and Electrical Insulation*, doi: 10.1109/TDEI.2025.3542343.

[19] H. Shadfar and H. R. Izadfar, "Frequency Response Analysis: An Overview of the Measurement Process and Interpretation of Results for Fault Diagnosis and Location in Power Transformers," *International Journal of Industrial Electronics Control and Optimization,* vol. 8, no. 2, pp. 149-163, 2025, doi: 10.22111/ieco.2024.49470.1603

[20] M. M. Mohsenzadeh, S. Hasanzadeh, H. R. Sezavar, and M. H. Samimi, "Flashover voltage and time prediction of polluted silicone rubber insulator based on artificial neural networks," *Electric Power Systems Research,* vol. 221, p. 109456, 2023, doi: 10.1016/j.epsr.2023.109456

[21] H. Sezavar, N. Fahimi, S. Hasanzadeh, and A. S. Akmal, "Risk assessment of contaminated composite insulators in pre-flashover conditions," *Electric Power Systems Research,* vol. 230, p. 110256, 2024, doi: 10.1016/j.epsr.2024.110256

[22] A. H. Alshalawi and F. S. Al-Ismail, "Partial Discharge Detection Based on Ultrasound Using Optimized Deep Learning Approach," in *IEEE Access*, vol. 12, pp. 5151-5162, 2024, doi: 10.1109/ACCESS.2024.3350555.

[23] S. Li *et al.*, "Partial Discharge Data Enhancement and Pattern Recognition Method Based on a CAE-ACGAN and ResNet," *Symmetry,* vol. 17, no. 1, p. 55, 2024, doi: 10.3390/sym17010055.

[24] M. Messaoudi, S. M. Kameli, S. S. Refaat, H. Abu-Rub and M. Trabelsi, "Deep Learning Based Corona Discharge Severity Classification for High Voltage Equipment," *IECON 2024 - 50th Annual Conference of the IEEE Industrial Electronics Society*, Chicago, IL, USA, 2024, pp. 1-5, doi: 10.1109/IECON55916.2024.10905579.

[25] H. R. Sezavar, N. Fahimi and A. A. Shayegani-Akmal, "An Improved Dynamic Multi-Arcs Modeling Approach for Pollution Flashover of Silicone Rubber Insulator," in *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 29, no. 1, pp. 77-85, Feb. 2022, doi: 10.1109/TDEI.2022.3146531.

[26] H. R. Sezavar and S. Hasanzadeh, "Integrating Weibull analysis and KAN-ODEs for enhanced flashover prediction in contaminated composite insulators," *Electric Power Systems Research,* vol. 244, p. 111584, 2025, doi: 10.1016/j.epsr.2025.111584

[27] S. F. Stefenon *et al.*, "Wavelet group method of data handling for fault prediction in electrical power insulators," *International Journal of Electrical Power & Energy Systems,* vol. 123, p. 106269, 2020, doi: 10.1016/j.ijepes.2020.106269.

[28] Z. Zhang, F. Yang, H. Zhang, C. Zhou, Y. Li, and H. Liu, "A probabilistic neural network assessment method for insulator pollution level based on discharge noise," *Measurement,* vol. 242, p. 115869, 2025, doi: 10.1016/j.measurement.2024.115869.

[29] Y. I. Jang, J. Y. Sim, J.-R. Yang, and N. K. Kwon, "The optimal selection of mother wavelet function and decomposition level for denoising of DCG signal," *Sensors,* vol. 21, no. 5, p. 1851, 2021, doi:10.3390/s21051851.

[30] H. Farajpanah, A. Adib, M. Lotfirad, H. Esmaeili-Gisavandani, M. M. Riyahi, and A. Zaerpour, "A novel application of waveform matching algorithm for improving monthly runoff forecasting using wavelet–ML models," *Journal of Hydroinformatics,* vol. 26, no. 7, pp. 1771-1789, 2024, doi: 10.2166/hydro.2024.128

[31] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," *Artificial Intelligence Review,* vol. 54, pp. 1937-1967, 2021, doi: 10.1007/s10462-020-09896-5.

[32] A. A. Salem *et al.*, "Classification of RTV-coated porcelain insulator condition under different profiles and levels of pollution," *Scientific Reports,* vol. 14, no. 1, p. 22759, 2024, doi: 10.1038/s41598-024-73520-7

[33] N. Fahimi, H. R. Sezavar and A. A. Shayegani-Akmal, "Flashover Prediction of Polymer Insulators Based on Dynamic Modeling of Pollution Layer Resistance Using ANN," in *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 30, no. 1, pp. 122-130, Feb. 2023, doi: 10.1109/TDEI.2022.3207452.

Hamid Reza Sezavar was born in Qom, Iran, in 1991. He received a B.Sc. degree from the Sharif University of Technology, Tehran, Iran, in 2013 and a M.Sc. in electrical engineering from the University of Tehran, Tehran, Iran, in 2015. He then received his PhD in High Voltage from the University of Tehran, Tehran, Iran, in 2022. He is currently working toward an assistant professor position at Qom University of Technology. His principal research interests are High voltage engineering, outdoor insulators, Electrical discharge, and AI algorithms.

Saeed Hasanzadeh was born in Shirvan, Iran, in 1981. He earned his B.Sc. in electrical engineering from Shahroud University of Technology in 2003, followed by his MSc. and Ph.D. from the University of Tehran in 2006 and 2012, respectively. His MSc. thesis focused on High Voltage Engineering, while his Ph.D. dissertation explored Wireless Power Transfer (WPT). In 2013, he joined Qom University of Technology as an Assistant Professor in the Department of Electrical and Computer Engineering, rising to Associate Professor in 2022. He served as the department's Dean from 2018 to 2023. Dr. Hasanzadeh is an editorial board member of the Power Electronics Society of Iran (PELSI), a Technical Program Committee member of the IEEE PEDSTC Conference, and the Scientific Chair of ICREDG2025. He received Qom University of Technology's Top Research Prize in 2019 and 2023 and was recognized as an Outstanding Lecturer in 2020 and 2022. In 2023, he was named Qom province's top innovator for developing two innovative products in partial discharge detection. His research interests include power electronics, electrical machines, wireless power transfer, and high voltage engineering.