Feature Extraction and Classification of Partial Discharge Signals in C_4F_7N-Based Gas Insulated Systems: A Time-Domain and Machine Learning Approach

Rahmat Ullah, Alistair Reid, Manu Haddad, Mini Nambiar, Peter Taddei, and Matthew Barnett

***Abstract*— In the quest for an environmentally friendly alternative to SF_6 gas to reduce carbon emissions, the C_4F_7N gas mixture ($C_4F_7N/O_2/CO_2$) emerged as a good candidate due to its lower environmental hazards and good dielectric strength. However, these new gas mixtures bring transition-related challenges due to their altered partial discharge (PD) patterns and dynamics, which complicate reliable PD detection and classification in C_4F_7N-insulated systems. Manufacturing or assembly flaws in gas-insulated equipment can enhance local electric fields, triggering PDs. Identifying PD sources is critical for effective online monitoring of high-voltage equipment, enabling early detection of insulation degradation and preventing potential failures. Historically, phase resolved partial discharge patterns employed for defect identification. However, challenges arise when multiple PD sources active simultaneously, as their patterns may partially overlap, leading to misunderstanding in the identification process. In this research, a machine learning algorithm is developed to classify PD sources using statistical and probabilistic analysis of time-domain parameters of PD pulses. Several PD source topologies are designed including protrusion, metallic particles, and floating electrode arrangements. Various time-domain parameters of PD pulses, along with their peak amplitude, are measured. The model employs Weibull distribution analysis, statistical variables such as kurtosis and skewness, and machine learning approaches to effectively distinguish between various types of PDs. Achieving approximately 98 % accuracy demonstrates the model's effectiveness in classifying simultaneous PD sources. Moreover, the study confirms that time-domain statistical feature analysis significantly improves the reliability and safety of electrical systems using C_4F_7N gas mixtures.**

***Index Terms*— C_4F_7N, Defect classification, Time-domain feature extraction, Weibull distribution, Condition monitoring, Statistical features.**

I. INTRODUCTION

IN recent years, power utilities have been actively sought to reduce their carbon footprint by substituting SF_6 with eco-friendly gases in high voltage equipment such as gas-insulated switchgear (GIS) due to its high GWP of 22800 times that of CO_2 and high atmospheric lifetime [1]. Therefore, both the Kyoto Protocol and the Paris Agreement classify SF6 as a greenhouse gas that significantly contributes to global warming [2, 3]. The global effort to reduce ecological damage, particularly by reducing greenhouse gases, has an impact on this change. Although SF_6 is effective for insulation, it raises important environmental issues because of its GWP. As a result, it has become crucial to adopt more environmentally friendly insulating gases. Many power utilities are interested in fluoronitrile gas mixtures due to their low GWP and good dielectric strength. However, accurate identification and classification of defects in systems using alternative gases is critical to ensuring their reliability and safety. This study will examine the PD analysis of the $C_4F_7N/O_2/CO_2$ combination, which has a ratio of 5:13:82. This mixture ratio has been selected because it is representative of common ratios presently used in in-service equipment in the UK and elsewhere.

A. Defect and discharges in GIS/GIE

Defects and faults in dielectrics can lead to nonlinear distribution of electric field within the insulation, which may initiate localised discharges in high magnitude locations. These defects can occur during manufacturing or maintenance [4]. Typical internal defects in GIS/gas-insulated equipment (GIE) include HV conductor protrusions, free metal particles and floating potentials (see Fig. 1). Free-metal particles are a leading cause of insulation failure in gas insulated systems. They can be generated during assembly, installation, or operation when metal components rub against one another. Owing to their small size, these metal particles are influenced by electric-field forces, causing them to move and oscillate within the gas-insulated environment. If the metal particles move across a wide enough range, they can form conductive paths or arcs between the HV conductor and the enclosure, potentially causing serious damage to the systems. Metal protrusion defects in GIE refer to sharp metallic projections, typically on the HV conductor, formed during manufacturing, installation, or operation. These fixed protrusions concentrate the electric field at their tips, triggering PD under high-voltage

(Corresponding author: Rahmat Ullah). Rahmat Ullah, Alistair Reid and Manu Haddad, are with the Advanced High Voltage Engineering Research Centre, Cardiff University, Cardiff, UK

Mini Nambiar, Peter Taddei, and Matthew Barnett are with the SSEN Transmission, UK.

stress. The sharp tips of protrusions distort and amplify the local electric field, leading to stable PDs at rated voltage. Under overvoltage conditions, this field enhancement can trigger full breakdown and cause severe GIE faults. Floating-potential defects occur when components lose proper electrical contact, becoming isolated and electrically 'floating.' This alters the local electric field and can produce high-amplitude PDs. Every PD source produces its own unique pattern of discharges. Employing automated techniques that detect PD sources by identifying these patterns may prevent the failure of the system and help with online condition monitoring of equipment.

B. PD analysis techniques

In the last three decades, different methods have been employed for the automated recognition of these PD patterns [5-7]. Phase-resolved partial discharge patterns (PRPD) are considered an effective technique for analysing these PD patterns at different points of the AC cycle. Identifying multiple sources of PD activity within insulation is often the main goal in practical scenarios. When multiple PD sources are present, it can be challenging to distinguish and identify their overlapping patterns. Signals from multiple sources and interference will further complicate the situation [8]. Hence, the development of more advanced techniques to be applied to enhance the identification and measurement of the PDs generated by different sources could contribute to more accurate identification of the severity and type of defect in these cases. An alternative method involves investigating the time-domain parameters of a PD pulses. This technique assumes that pulses that originate from similar PD sources will contain similar features. This time domain parameter analysis can be helpful in categorising when multiple sources occur simultaneously.

Recently, researchers have employed both new and well-established machine learning algorithms to analyse PD measurements [9]. Several algorithms have been implemented and shown promising results in accurately identifying and locating faults [9–11]. Artificial neural networks have shown remarkable success in classifying patterns in PRPD data and recognising their statistical characteristics [12]. In addition, some authors [13] also utilised other methods of dimensionality reduction, such as principal component analysis (PCA) to enhance the classification results. Moreover, several techniques have been employed for PD classification in transformers and GIS insulated with SF_6 [14–16]. However, studies on the classification of PD defects in C_4F_7N-based gas-insulated systems are limited, despite being essential for the transition from SF_6 to environmentally friendly alternatives.

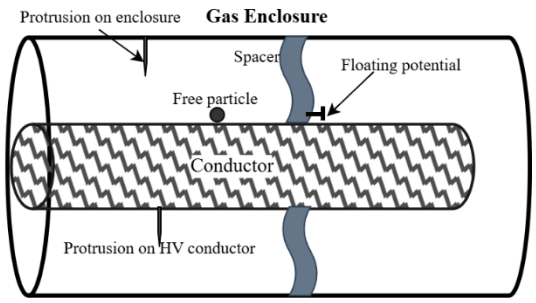

Fig. 1. Gas insulated system various PD defects.

Therefore, this study aims to develop an algorithm capable of accurately classifying PD defects in C_4F_7N mixture-based gas-insulated systems using time-domain and statistical features of the PD pulses. To achieve this, the Random Forest technique was chosen due to its effectiveness in handling large, nonlinear datasets with high accuracy while minimizing the risk of overfitting. In [17], the authors have applied several techniques for the classification of the acoustic partial discharge of high-voltage busbar systems. Out of all the techniques, random forest classifier consistently outperforms all other classifiers in terms of accuracy when dealing with noisy inputs. Various time-domain parameters—such as rise time, fall time, pulse width, rise and fall slew rates, and peak amplitude—were calculated from the acquired PD pulses for all defect types. After that using the likelihood technique, the 2-parameter Weibull distribution including scale and shape are calculated. The statistical fingerprints, including skewness and kurtosis, of all the PDFs are calculated for all PD pulses. All these parameters were then used to train the algorithm. The trained model was subsequently tested with a set of test data to evaluate its classification accuracy.

II. EXPERIMENTAL ARRANGEMENTS FOR PARTIAL DISCHARGE MEASUREMENTS

Fig. 2 depicts the experimental setup developed to evaluate PD within a high-voltage test vessel. The core component of the setup involves the HV power supply, which supplies the necessary voltage to the test vessel through HV bushing which is PD free up to 38 kV. The resistor and capacitor are utilised to create a low pass filter that effectively mitigates noise originating from the amplifier. A high-frequency current transformer (HFCT) having 5 V/A sensitivity was used to capture the high-frequency signals associated with PD events. A parallel resistor-capacitor voltage divider was used to monitor the test voltage. The test defects were installed in a stainless-steel chamber having a pressure rating up to 20 bar absolute. The data presented in this work was measured at 8 bar absolute pressure, which is the rated pressure in high-voltage gas-insulated systems.

Obtaining data is a crucial element of the setup which involves extracting a phase reference signal from the power supply to synchronise the PD measurements with the AC voltage supply cycle. In addition to the PD detection with the HFCT, the signals are recorded with an oscilloscope for initial capture and display of the data along with voltage and phase references simultaneously. The oscilloscope is connected to a national instruments data acquisition (NI-DAQ) system. This enables a thorough

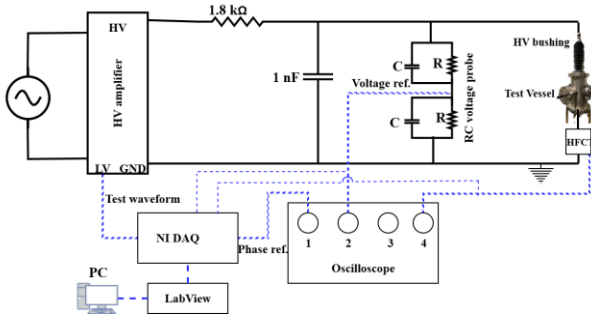

Fig. 2. Experimental arrangements for PD measurements.

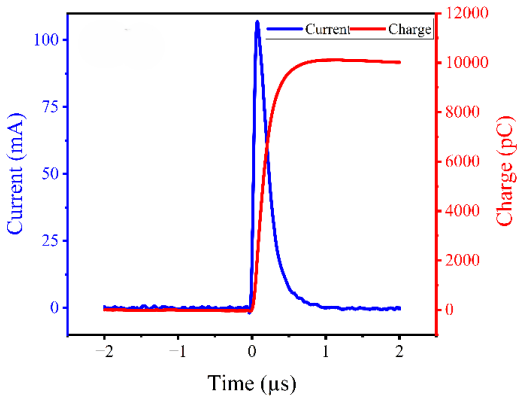

Fig. 3. Charge injector for calibration.

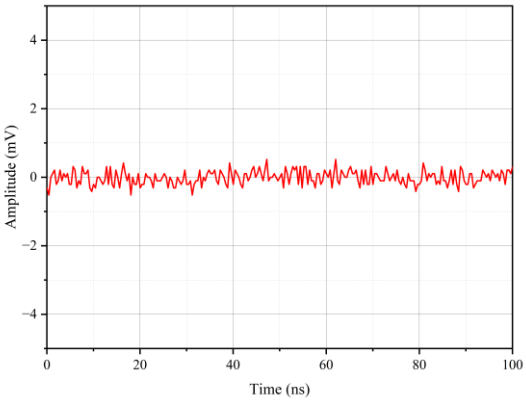

Fig. 4. Noise level measured at the lab.

examination of the PD signals in relation to the voltage phase, which improves the diagnostic capabilities for PD analysis. Before starting the actual experimentation, calibration of the setup was achieved by injecting different levels of charges into the terminal and then capturing them on the oscilloscope. The setup can detect a wide range of charge levels, starting from as low as 5 pC, demonstrating the high accuracy of the testing setup and one of the charge injectors is shown in Fig. 3. The noise level was measured and found to be less than 2 mV peak-to-peak as displayed in Fig. 4. The oscilloscope's trigger level was adjusted to a value slightly higher than the noise level to ensure that only PD pulses activated the equipment and reduced interference from noise. The voltage

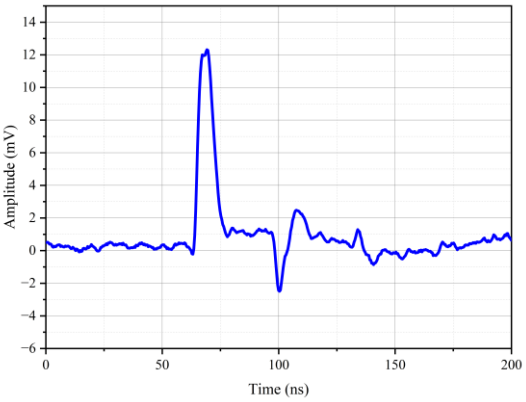

Fig. 5. Typical PD pulse measured by HFCT.

was gradually raised from zero at a rate of 1 kV/s. After that, the voltage was increased and the PD signal was measured, which was clearly differentiated from the noise signal (as displayed in Fig. 5).

To assess the effectiveness of the adopted method, it is essential to replicate realistic gas insulated system conditions within a controlled laboratory setting. It allows for the creation of accurate models that represent protrusions, free metallic particles and floating electrodes. The following section explores the complexities of the design.

The initial setup showcases a protrusion model featuring a sharp needle-like structure that is 25 μm thick, extends 10 mm from a smooth conductor surface and spans 120 mm in length as shown in Fig. 6(a). The protrusion's sharpness and length represent an idealised defect that may arise due to manufacturing flaws or material deteriorations that cause abnormalities on conductor surfaces. The protrusions are crucial stress areas that intensify electric fields and are very prone to causing partial discharges, particularly under high-voltage conditions. The precise geometry of the protrusion is essential for accurately analysing the field enhancement and the threshold voltages required for discharge initiation.

The second setup shows a metallic particle having a diameter of 4 mm positioned above a concave-shaped conductor labelled HV, causing an irregular field gradient, as shown in Fig. 6(b). This configuration exemplifies situations in which loose conducting elements, either from mechanical wear or manufacturing remnants, become conductive pollutants inside the GIS. Particles subjected to electrostatic forces may exhibit intermittent discharges as they travel inside an electric field.

The third configuration shows a spherical floating electrode placed centrally over a flat conductor with a defined 1 mm gap, as shown in Fig. 6(c). The floating electrode in the GIS context is anticipated to cause a notable distortion in the electric field distribution. When an electrode is in a floating condition, it is electrically isolated, enabling the buildup of charge that can result in localised field increases. These may raise the likelihood of PD initiation, making it a relevant model for researching discharge starting and spreading because of the floating electrode defects within gas-insulated switchgear.

IV. RESULTS AND DISCUSSION

A. Time-Domain Feature Analysis in PD Detection

PD pulses originating from the same source generally exhibit similar waveform shapes, which facilitates efficient waveform analysis. Time domain features such as rise time, fall time, slew rates in rise and fall and pulse width were selected for the analysis because these parameters are conducive to classifying defects from multiple PD source configurations. Fig. 7 displays details of these parameters for one of the PD pulses. The time domain parameters acquired for all the PD pulses as follows:

- **Rise time:** when the signal reached from 10% to 90% of the peak amplitude then that time interval is considered as rise time.
- **Fall time:** When the signal fall from 90% to 10% of peak amplitude then that interval is considered as fall time.

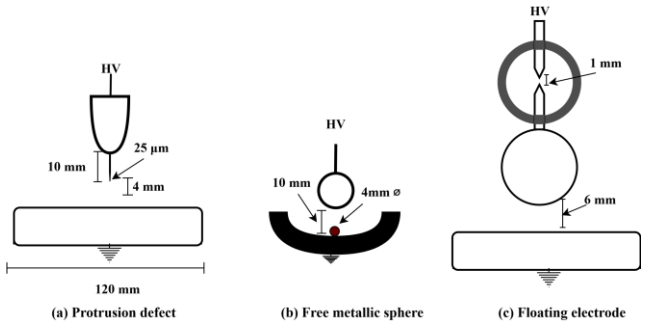

Fig. 6. PD defects (a) protrusion, (b) free metallic sphere and (c) floating electrode.

- **Pulse width** corresponds to the duration between the points where the pulse crosses a predefined threshold on the rising and falling edges.
- **Slew rate (rise and fall):** Slew rates indicate how quickly the pulse rises and falls, measured as the rate of change of the signal. It can be calculated by dividing the change in voltage by either the rise time or the fall time.

PD measurements were carried out for each of the PD sources, such as protrusion in the C_4F_7N mixture (C4P), floating electrode in the C_4F_7N mixture (C4F), free metallic electrode in the C_4F_7N mixture (C4M), and protrusion in the air (AP). For each PD source, 5000 segments of data were collected. After that, time domain parameters were calculated for all PD pulses, with the mean values shown in Table 1. Additionally, the peak amplitude of each pulse was measured, with average values reported in Table 1 for each PD source.

The data in Table 1 for different PD sources reveals distinct behaviour under various conditions. The rise and fall times for the protrusion in air are lower than those for the protrusion in the C_4F_7N mixture. Moreover, the lowest rise and fall times were observed for the free metallic electrode arrangement. On the other hand, slew rates and peak amplitude were noticed as important parameters for classification. The slew rates for floating electrodes are highest, followed by metallic electrodes and then protrusion in air and the C_4F_7N mixture, respectively. Also, the peak amplitude values are beneficial because each PD source produces an identical range of values. The highest amplitude was observed for the floating electrode, and the smallest one was noticed for the protrusion in the air. The different parameters provide specific details about the speed, duration, and intensity of discharges, allowing for an improved identification and differentiation of defects based on their electrical properties.

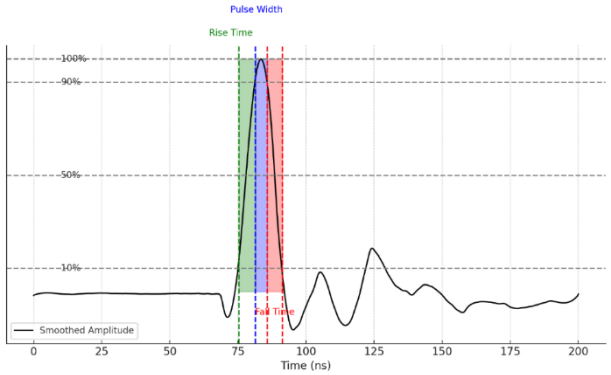

Fig. 7. Time domain parameters of a typical PD pulse.

TABLE I AVERAGE VALUE OF DIFFERENT PARAMETERS OF PD PULSES.

	Air-protrusion	C4-protrusion	C4-metallic	C4-floating
Rise time (ns)	6.05	7.24	5.60	17.13
Fall time (ns)	8.75	11.30	5.66	14.67
Pulse width (ns)	11.63	18.54	11.27	31.81
Slew rates rise (ns)	3.63	0.076	19.95	59.59
Slew rate fall (ns)	3.94	0.068	21.20	69.58
Peak Amplitude (mV)	22.75	52.77	117.66	1020.6

B. Statistical Diagnostics for Partial Discharge Identification

This section introduces the statistical techniques developed to identify the sources of measured partial discharges. The primary aim of using statistical diagnostics in this context is to detect and accurately locate PD sources in SF_6-alternative electrical insulation systems. Statistical models can play an important role in the classification problem in this study, and our focus is on the normal and two-parameter Weibull distributions. These distribution methods were shown to be effective when analysing different parameter variations in insulation failure. Hence, they were selected based on their

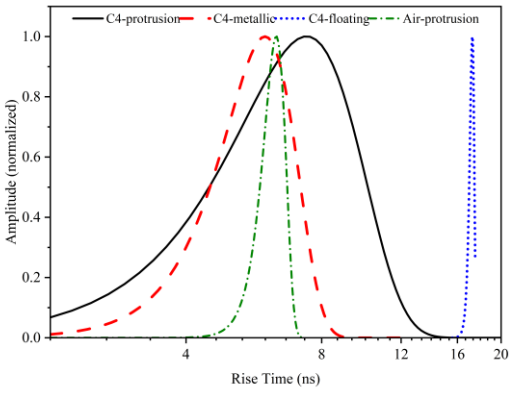

Fig. 8. PDF curves of rise time.

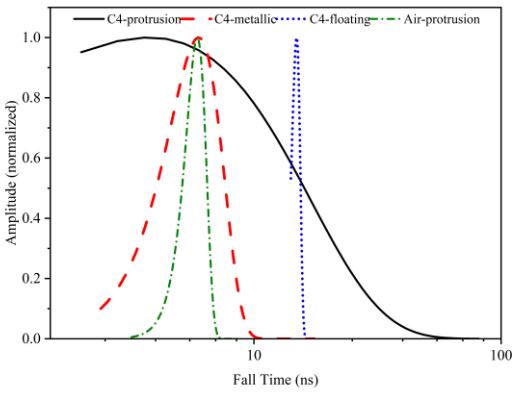

Fig. 9. PDF curves of fall time.

demonstrated effectiveness in comparable engineering situations.

1) 2-parameter Weibull distributions

The Weibull distribution is highly valuable because of its ability to effectively model data across a wide range of shapes due to its flexible parameters particularly the shape and scale parameters, as modelled in Equation (1). The outcome is heavily influenced by two key parameters: (a) the shape parameter represented by k, and (b) the scale parameter represented by λ. The shape parameter influences the shape of the PDF curve, while the scale parameter is responsible for stretching out the PDF curves. Using the likelihood method and two parameters, Equation (1) allows the calculation of the probability density function (PDF) curves.

$$f(x;\,\lambda,k)\left(\frac{k}{\lambda}\right)\left(\frac{x}{\lambda}\right)^{(k-1)} e^{\left(-\left(\frac{x}{\lambda}\right)^{k}\right)} \qquad (1)$$

Where x is the variable, λ (lambda) is the scale parameter, k is the shape parameter, $f(x; \lambda, k)$ is the value of the PDF at x. Note: The distribution is applicable for $x \geq 0$, $\lambda > 0$, and $k > 0$. The parameter k defines the distribution type, whereas the parameter λ enables adjustments along the x-axis.

Fig. 8 shows the PDF curves depicting the rise time values from various PD sources, comparing the variations in air and C_4F_7N mixture. Fig. 8 indicates that different sources of PD produce PDF curves with varying rise time distributions. In comparison to other PD sources, the rise time of the floating electrode is significantly higher. Also, the PD pulses, that are caused by fixed protrusion defects in the C_4F_7N mixture distribution have shorter rise times than those in air. This could be due to the presence of high-density C_4F_7N gas in the mixture, which can quench the discharge much faster than air. Although the PDF curves in the figure show some overlaps, it is possible to distinguish between the protrusion in air and C_4F_7N using the rise time PDF curves. Additionally, the PDF for the floating electrode is remarkably similar. On the other hand, the graph shows that there is a significant overlap between the rise time distributions of discharges caused by a protrusion in air and moving metallic particles in a C_4F_7N mixture. According to the rise time PDF distribution curves, it appears that this could not be enough to fully differentiate between multiple PD sources and requires some additional parameters. Fig. 9 and Fig. 10 display the fall-time and pulse width PDF curves for all the PD sources. It is difficult to differentiate between these PD sources from the fall-time and pulse width PDF curves because all the curves overlap each other.

Fig. 11 and Fig. 12 display the PDF curves illustrating the slew rates in the rise and fall of the curves. The PDF curves clearly differentiate between the slew rates of each PD source. Thus, the slew rates play a crucial role in categorising various faults in the C_4F_7N gas insulated system, as well as in detecting corona discharge in air. Fig. 13 showcases the PDF curves illustrating the peak amplitude of the PD pulse for all PD sources. PD pulses of varying peak amplitudes are produced by all PD sources. For the case of protrusion in air, it is interesting to note that it has a lower peak amplitude compared to other sources of PD. On the other hand, the floating electrode

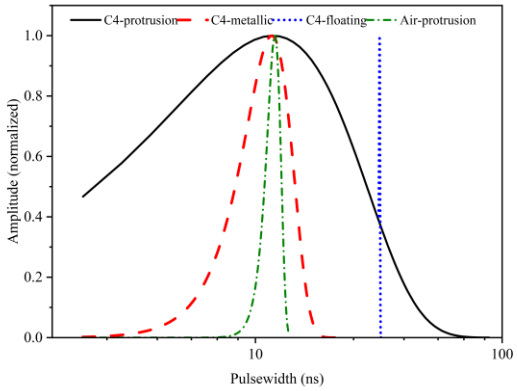

Fig. 10. PDF curves of pulse width.

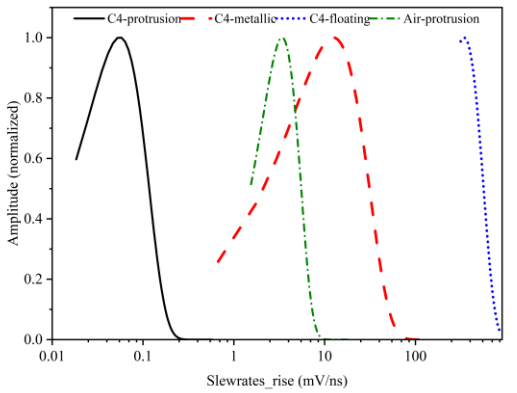

Fig. 11. PDF curves of slew rates in rise.

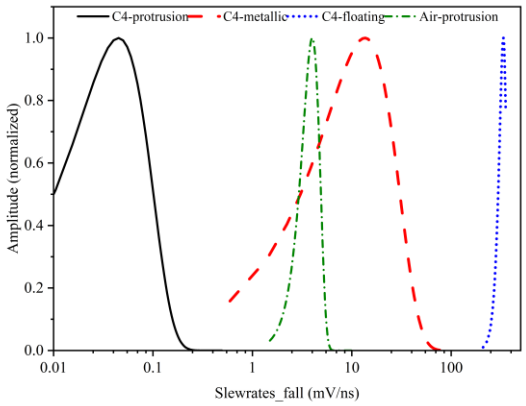

Fig. 12. PDF curves of slew rates in fall.

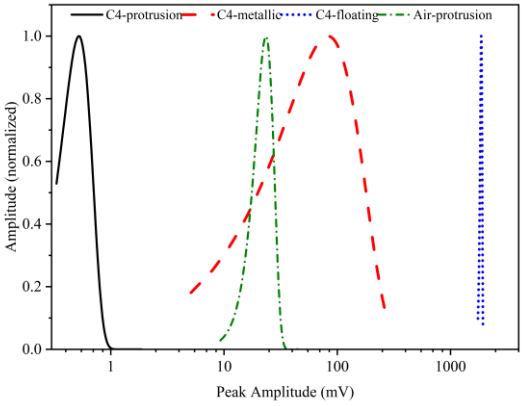

Fig. 13. PDF curves of peak amplitude

TABLE 2. SKEWNESS AND KURTOSIS OF ALL THE PD SOURCES.

	RT		FT		PW		SR_R		SR_F		MA	
	S	K	S	K	S	K	S	K	S	K	S	K
C4-protrusion	2.76	36.70	2.78	10.75	2.52	10.06	5.89	44.36	4.13	36.06	4.91	42.77
C4-metallic	1.44	23.84	3.65	30.73	-0.93	12.77	1.50	10.07	0.50	3.94	0.08	1.69
C4-floating	-2.74	10.22	2.13	7.13	1.83	7.92	0.39	2.53	-0.37	2.61	0.35	2.36
Air-protrusion	-3.83	23.17	0.77	13.24	-4.03	19.90	6.44	52.40	1.43	16.07	1.13	9.98

TABLE 3. SCALE AND SHAPE PARAMETERS OF THE PD SOURCES.

	RT		FT		PW		SLW_R		SLW_F		MA	
	k	λ	k	λ	k	λ	K(e8)	λ	k	λ	K(e7)	λ
C4-protrusion	8.10	3.60	13.20	1.25	20.89	1.63	0.85	1.82	0.74	1.72	0.05	3.69
C4-metallic	6.18	5.84	6.33	4.35	12.18	4.94	23.1	1.59	22.4	1.72	13.19	1.77
C4-floating	17.27	89.94	14.90	26.24	31.87	202.92	61.9	12.13	71.89	16.02	105.1	13.16
Air-protrusion	6.36	18.06	5.95	10.90	12.03	15.66	4.1	2.50	4.15	5.3	2.42	5.73

arrangement stands out with the highest peak amplitude. This feature is crucial, as all the PD sources generate distinct PDF curves that can be easily distinguished from one another, as depicted in Fig. 13.

2) Probability Distribution Metrics

The different metrics of probability distribution can be calculated and then used for evaluation and classification purposes. Important parameters of the probability distribution are skewness (S) and kurtosis (K). Gaining insight into the absence of symmetry around the mean of a PDF distribution can be accomplished by assessing its skewness. On the other hand, kurtosis measure the sharpness of the PDF distribution. Skewness and kurtosis can be found from Equations (2) and (3).

$$Skewness = \frac{E((x-\mu)^3)}{N\sigma^3} \qquad (2)$$

$$Kurtosis = \frac{E((x-\mu)^4)}{N\sigma^4} \qquad (3)$$

Where, μ represents the mean of the data, N represents the number of samples, and x represents the feature value. Both equations are normalised by σ^3 and σ^4. For all the data, skewness and kurtosis are measured as shown in Table 2. Positive skewness indicates a distribution of most data on the right side of the mean, while negative skewness indicates a concentration of data on the left side of the mean. Positive kurtosis leads to a sharper distribution of the data points, and negative kurtosis shows a flatter distribution. Also, the statistical moments λ and k have been measured and displayed in the table. Tables 2 and 3 display the total 24 statistical moments measured for all parameters. Using these statistical moments, an algorithm was developed in this work that was then used for the classification of multiple PD sources.

3) Random Forest

Given the extensive range of selected features and their complex categorization, it might be beneficial to employ machine learning for classification purposes. By analysing different time domain parameters of a PD pulse, it is possible to distinguish four PD sources in a gas insulated systems using random forest (RF). Leo Breiman [18] developed the Random Forest (RF) technique in 2001 which is a machine learning methodology that combines multiple decision trees to form an ensemble model. The algorithm has two built-in randomization mechanisms that make RF strong at generalisation and resistant to interference. By using these two techniques, RF assesses the importance of features and then deploys feature selection. It is important to consider three parameters: (a) the size of the feature subset, (b) the criteria for splitting nodes, and (c) the number of trees that have a significant impact on the performance of RF. Empirical evidence suggests that 500 trees are sufficient to meet the categorization criterion [19], and more than 500 would only increase the computational time. The RF technique is efficient when dealing with large datasets with varying dimensions. It also solves the problem of overfitting, which is a major issue in the decision tree technique. The RF builds the tree through bootstrap sampling of the data, a technique that fosters diverse learning and bolsters the tree's resilience. The RF concludes the decision using a majority voting mechanism. Also, it avoids the extensive pre-processing of the diverse and complex data. This study's methodology significantly enhances prediction accuracy and offers valuable insights into the significance of various attributes in discharge source prediction. A technique exploiting this method is developed in this work as detailed below.

4) Algorithm for classification

The step-by-step detail of the developed algorithm is illustrated in Fig. 14. Initially, the data is pre-processed by labelling and making a combined data set. Then, the data is split into two parts, one for training purposes and the other for testing purposes. Both subsets undergo a filtering process to minimise noise and enhance signal features. Various statistical features were extracted and modelled including rise time, fall time, pulse width, slew rates (fall and rise) and peak amplitude Probabilistic analysis was undertaken by finding the PDF distribution using the likelihood method. This enables the construction of PD fingerprints defined by the parameters λ and k. Also, the statistical analysis of a PDF distribution was done

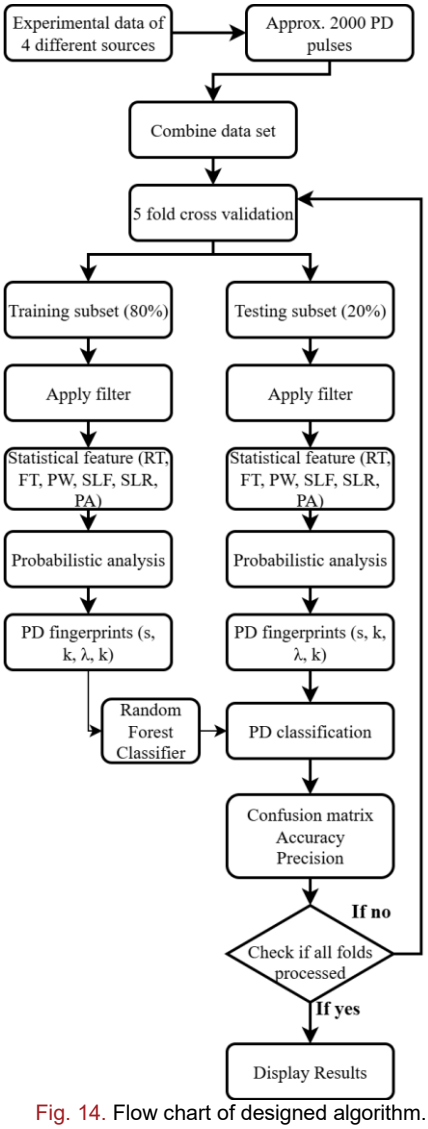

Fig. 14. Flow chart of designed algorithm.

based on its skewness and kurtosis.

The classification of PD defect types was achieved by training a Random Forest classifier using time-domain PD pulse parameters along with their statistical features. Moreover, for each feature the Weibull distribution parameters such as scale (λ) and shape (k) were estimated to analyse discharge pulse dynamics, and skewness and kurtosis were calculated to evaluate how closely the behaviour matched a Gaussian distribution. The Random Forest model configured with 100 trees and validated using 5-fold cross validation used these feature parameters to establish the decision boundaries between classes. Additionally, when evaluating feature importance using the Gini index which is calculated by using equation 4, slew-rate rise, and pulse width emerge as key parameters for distinguishing between types of PD defects.

Table 4 shows the discriminative power of the five time-domain parameters based on the Weibull parameters and statistical parameters. The Δλ and Δk showing the change in the maximum and minimum values for each feature among all the defects. It can be observed that the pulse width having high Δλ (range: 1.63-202.92) makes it important for identifying the

TABLE 4. DISCRIMINATIVE POWER OFDIFFERENT FEATURES.

Features	Δλ	Δk	Max |S|	Max |K|	Score	Gini Index
SR_R	10.31	61.05	5.89	44.36	0.89	0.32
PW	201.2	19.69	4.03	19.90	0.85	0.28
RT	86.34	9.17	3.83	36.70	0.79	0.19
SR_F	14.30	71.15	4.13	36.06	0.72	0.15
FT	24.99	8.57	3.65	30.73	0.68	0.12

floating defects. Moreover, slew rate rise Δk have the largest value makes it critical detecting protrusion defects. On the other hand, fall time has a narrow range of λ (1.25-26.24) and k (6.33-14.90) and overlapping values across all defects which makes separability difficult. To rank the importance of each feature, the classification separation score has been calculated by using equation 2. The slew rate rise score of 0.89 dominated by Δk and skewness makes it ideal for identification of protrusion whereas the pulse width have the second largest score makes it critical for floating defects. These rankings can be validated through Gini index which show how important each feature contributes to the accuracy of the algorithm.

$$Score = 0.25 \times \frac{\Delta\lambda}{\max(\Delta\lambda)} + 0.25 \times \frac{\Delta k}{\max(\Delta k)} + 0.30 \times \frac{|S|}{\max(|S|)} + 0.25 \times \frac{K}{\max(K)} \quad (1)$$

The output performance of the designed algorithm was measured by employing different parameters, such as accuracy and the confusion matrix. The process is repeated for each fold of cross-validation. Once all the folds are completed, the results are compiled and presented in terms of the confusion matrix for each fold. The confusion matrix demonstrated the classification of four types of PD defects in each row. The green box in Fig. 15 highlights the true classification for each defect, while the red box indicates the missing classification. From the confusion matrix for each fold, it is evident that the miss classification number is significantly lower than the true classification, indicating the high accuracy of the developed algorithm.

Also, the developed algorithm's accuracy has been checked by varying the number of features. As the number of features increases, the classification accuracy of the algorithm also increases, as shown in Fig. 16. Using a limited number of features, typically less than 8–10, significantly reduces the accuracy of initial identification for PD classification. By incorporating a larger number of input features, the accuracy of the four PD types can be significantly improved, reaching an impressive 99%. It is evident that the chosen features have a high level of representativeness.

V. Conclusion

This paper demonstrates the results of classification of concurrent PD sources in a C_4F_7N gas mixture using a machine learning algorithm based on statistical PD pulse features. For this purpose, PD data consisting of 2000 PD pulses were captured from four tests using different PD sources. Multiple statistical features were extracted from the captured data. Also, the developed approach integrated 2-parameter Weibull distribution of time domain parameters such as rise time, fall time, pulse

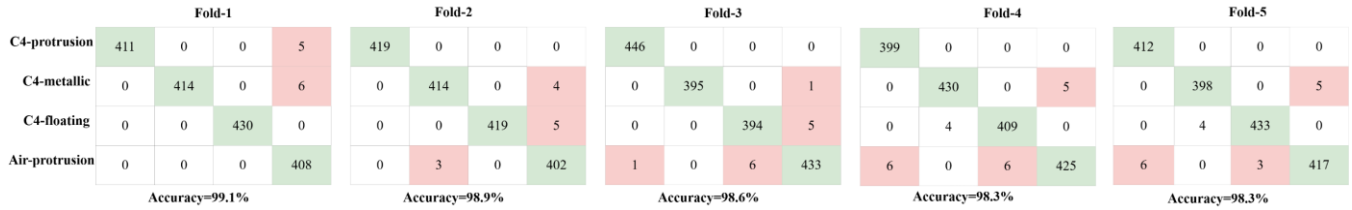

Fig. 15. Confusion matrix for fold 1-5.

width, slew rate rise and slew rate fall and peak amplitude, along with statistical metrics such as skewness and kurtosis for the PD defects. These parameters were fed to a Random Forest classifier for training. This algorithm enhances the classification process, delivering high accuracy of more than 98% which is also accompanied by significant time savings, hence its potential suitability for condition monitoring.

In addition, the implementation of this methodology has the potential to supply network operators and manufacturers of diagnostic equipment with data that may improve the reliability of electrical systems that use C_4F_7N, a more environmentally friendly alternative to SF_6.

Future work will corroborate the reported results with online data from electric utilities and expand the classification methods

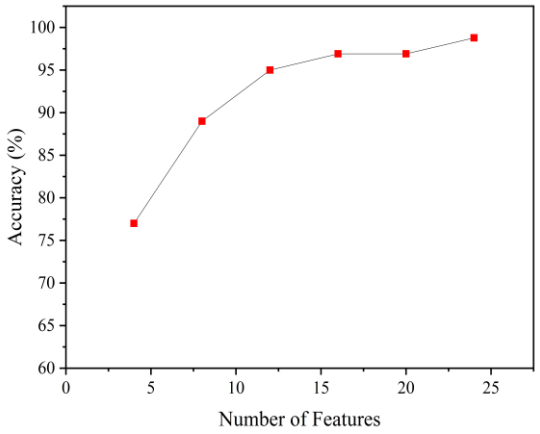

Fig. 16. Relationship between accuracy and number of

using advanced technique and measurement systems.

REFERENCES

[1] IPCC, “Climate Change 2013: The Physical Science Basis. Working, Group I: Contribution to the Fifth Assessment Report on Intergovernmental Panel on Climate Change”, 2013.

[2] United Nations. Kyoto Protocol—United Nations Framework Convention on Climate Change. 1998.

[3] United Nations. Paris Agreement—United Nations Framework Convention on Climate Change. 2015.

[4] Kessler, Ole. "The importance of partial discharge testing: PD testing has proven to be a very reliable method for detecting defects in the insulation system of electrical equipment and for assessing the risk of failure." *IEEE Power and Energy Magazine* 18, no. 2 62-65, 2020.

[5] Sahoo, N. C., M. M. A. Salama, and R. Bartnikas. "Trends in partial discharge pattern classification: a survey." *IEEE Transactions on Dielectrics and Electrical Insulation* 12, no. 2 248-264., 2005.

[6] Raymond, Wong Jee Keen, Hazlee Azil Illias, and Hazlie Mokhlis. "Partial discharge classifications: Review of recent progress." *Measurement* 68 164-181, 2015.

[7] Gulski, E., and F. H. Kreuger. "Computer-aided recognition of discharge sources." *IEEE Transactions on Electrical Insulation* 27, no. 1 82-92, 1992.

[8] Ma, Hui, Jeffery C. Chan, Tapan K. Saha, and Chandima Ekanayake. "Pattern recognition techniques and their applications for automatic classification of artificial partial discharge sources." *IEEE Transactions on Dielectrics and Electrical Insulation* 20, no. 2 468-478, 2013.

[9] Lu, Shibo, Hua Chai, Animesh Sahoo, and B. T. Phung. "Condition monitoring based on partial discharge diagnostics using machine learning methods: A comprehensive state-of-the-art review." *IEEE Transactions on Dielectrics and Electrical Insulation* 27, no. 6 1861-1888, 2020.

[10] Soltani, Amir Abbas, and Ayman El-Hag. "Denoising of radio frequency partial discharge signals using artificial neural network." *Energies* 12, no. 18 3485, 2019.

[11] Iorkyase, Ephraim T., Christos Tachtatzis, Pavlos Lazaridis, Ian A. Glover, and Robert C. Atkinson. "Low-complexity wireless sensor system for partial discharge localisation." *IET Wireless Sensor Systems* 9, no. 3 158-165, 2019.

[12] Mas’ud, Abdullahi Abubakar, Brian G. Stewart, and Scott G. McMeekin. "An investigative study into the sensitivity of different partial discharge φ-qn pattern resolution sizes on statistical neural network pattern classification." *Measurement* 92 497-507, 2016.

[13] Khan, Yasin. "Partial discharge pattern analysis using PCA and back-propagation artificial neural network for the estimation of size and position of metallic particle adhering to spacer in GIS." *Electrical Engineering* 98 29-42, 2016.

[14] Hussein, Ramy, Khaled Bashir Shaban, and Ayman H. El-Hag. "Robust feature extraction and classification of acoustic partial discharge signals corrupted with noise." *IEEE Transactions on Instrumentation and measurement* 66, no. 3: 405-413, 2017

[15] Woon, Wei Lee, Ayman El-Hag, and Mustafa Harbaji. "Machine learning techniques for robust classification of partial discharges in oil–paper insulation systems." *IET Science, Measurement & Technology* 10, no. 3: 221-227, 2016.

[16] Desai, BM Ashwin, Ramanujam Sarathi, Joe Xavier, and A. Senugupta. "Partial discharge source classification using time-frequency transformation." In *2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIS)*, pp. 362-366. IEEE, 2018.

[17] Hussein, Ramy, Khaled Bashir Shaban, and Ayman H. El-Hag. "Robust feature extraction and classification of acoustic partial discharge signals corrupted with noise." *IEEE Transactions on Instrumentation and measurement* 66, no. 3: 405-413, 2017.

[18] L. Breiman, “Random forests,” *Mach. Learn*., vol. 45, no. 1, pp. 5–32, 2001.

[19] Hasan, Md Al Mehedi, Mohammed Nasser, Shamim Ahmad, and Khademul Islam Molla. "Feature selection for intrusion detection using random forest." *Journal of information security* 7, no. 3: 129-140, 2016.