Partial Discharge Source Classification Using Pattern Recognition Algorithms

by

Hamed Janani

A Thesis Submitted to the faculty of Graduate Studies of
The University of Manitoba
in Partial Fulfilment of the Requirements for the Degree of
Doctor of Philosophy

Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba, Canada

Abstract

Design, development, and testing of a comprehensive and automated classification system for single and multiple partial discharge (PD) source identification based on the relationship between the variation of phase resolved partial discharge (PRPD) patterns and the sources of PD is proposed. The proposed system consists of feature extraction methods and classifier algorithms that are implemented for recognition of partial discharge patterns. Once the PRPD patterns are recorded, features generation algorithms are applied on the collected data. For single PD source identification, twelve high performance, applicable feature extraction techniques on PRPD patterns are employed to extract features. In order to present a comprehensive classification system, 10 well-known algorithms for the classification of PD sources have then been used. To evaluate the performance of the classification system, three laboratory test setups are designed and built to simulate various types of PD activities. The first test setup includes test cells which are designed to model common sources of PD in air, oil, and SF_6. Using this setup, the application of automated classification system on different sources of PD in different HV insulation media is investigated. The second and third test setups are designed to test the classification system on identification of different sources of PD in oil-immersed insulation and power transformer cellulose insulation under both electrical and thermal stresses, respectively. In many practical situations, the interest lies in the identification of multiple, simultaneously activated PD sources in insulation. Multi-source PDs sometimes results in partially overlapped patterns, which makes them hard to be identified by single source identification techniques. To further enhance the proposed classification system, a novel algorithm to identify Multi-source PDs is developed and appended to the system. To evaluate the performance of this

algorithm, a number of multi-source PD models have been designed. The overall results show that the classification system is well able to identify the single and multi-source of partial discharges. More importantly, this identification system is able to assign a "degree of membership" to each PRPD pattern, besides assigning a class label to it. This enables probabilistic interpretation of a new PRPD pattern that is being classified and results in safer decision making based on the risk associated with different sources of PD. The results of this research is beneficial for the design of a solid basis for an automated, continuous 24/7 monitoring of equipment, which facilitates PD source identification in early stages and safe operation of HV apparatus.

Acknowledgements

I would like to extend my utmost gratitude to the many people who have helped along the way to make this research project possible and to bring it to its completion. First, I would like to express my sincerest appreciation to Dr. Behzad Kordi for providing me the opportunity to take part in his research work. I am so deeply grateful for his help, invaluable guidance, and support throughout this project and my entire program of study.

I would like to thank the committee members Dr. Mohammad Jafari Jozani, Dr. David Swatek and Dr. Shesha H. Jayaram for their advice, guidance, and support through the completion of the project. Without their participation and input, the project could not have reached the level of fruition it has achieved.

Furthermore, many thanks to Nathan Jacob who assisted with the lab experiments at Manitoba Hydro High Voltage Test Facility and accompanied me when running some tests and conducting measurements.

I must also express my very profound gratitude to my parents; my dear mother Razieh, who always brightens my day with her positive energy, and for my reassuring father Ali, for making me find the strength to continue my research, and to the rest of my family for the many thoughts and words of encouragement that helped me get through challenging times.

Lastly, I want to thank my friends for their unfailing encouragement and support throughout my years of study, researching, and writing this thesis. Thank you all for all the wise words, your efforts and everything you did for me, I could not have done it without you.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Safe, stable, and reliable electric power systems rely on solid, liquid, and gaseous insulation materials to isolate energized components from other components and the ground. These insulation materials experience large electrical stresses during operation, especially in high voltage (HV) environments where the stress causes nanoscale molecular defects (ageing). These defects, in turn, become concentration points for the electrical stress gradually resulting in a proliferation of defects and the creation of micron-scale defects in the material. Once the defect achieves a critical size, the electric field can cause small, local breakdowns known as partial discharges (PD) to occur within the defect. This threshold also marks the beginning of a much faster deterioration of the material condition during which PD activity is associated with increasing rates of degradation leading to catastrophic failure (breakdown) [1].

Monitoring partial discharges, as a symptom of insulation deterioration, can be used to improve the reliability of HV insulation. Early detection of PD activities prevents costly failures of electrical equipment. The techniques employed for PD detection are based on chemical, acoustic, optical, electrical, and Ultra High Frequency (UHF) measurements. The electrical measurement of PD in

high voltage AC systems are widely used and are the focus of this thesis. In general, the motivation of this PhD thesis is the need of electric power system industry to diagnose defective insulation systems and also to classify the type of defects in early stages. Such measurement and classification must have the capability to be performed in both online or off line modes in order to prevent costly failures due to HV apparatus breakdowns.

One important application of PD measurements is in the identification[1] of the source of PD [4]. Partial discharge can be used to perform online condition assessment monitoring of HV insulators and evaluate the reliability of HV insulation systems [4]. In AC systems, the phase-resolved partial discharge (PRPD) pattern, which visualizes the occurrence of PD activities in reference to the phase of AC voltage, has been a valuable diagnosis tool. In a PRPD pattern, two important parameters, both in reference to phase angles, are the discharge magnitude and discharge rate. They form a bivariate distribution where each of the discharge magnitude and discharge rate can be separately analyzed with reference to each other and to the phase angle of the AC source [5]. Each type of insulation defect has its own discharge mechanism and features, and as such, it leads to the generation of a unique discharge pattern. Visual inspection of PRPD patterns by human experts has been one of the major partial discharge analysis approaches. However, in recent years, due to the availability of high-speed data processors and well-developed statistical techniques in machine learning, automated identification of PD sources seems to be more achievable [4]. Reliable, automated, classification of PD sources enables online monitoring of high voltage (HV) apparatus more accurately and efficiently. In this way, the ability to identify defects in early stages can lead to safety augmentation of HV apparatus, such as transformers, electric machines, cables, and Gas-Insulated Switchgear (GIS).

Existence of a correlation between the nature of PD sources and their PRPD patterns, has been the motivation of designing a thorough automated feature extractor and pattern classifier system

[1]The terms identification and classification can be used interchangeably, however it is worth to mention based on their exact meaning, identification is the process of recognizing an unknown object and name it, and classification is the statistical analysis to assign an unknown object into a category of objects [2, 3].

for the application in the area of HV insulation monitoring.

1.2 Objectives

The main objective of this PhD thesis is to study the correlation which exists between the nature of partial discharge sources in different HV insulation media and their corresponding PRPD patterns. Based on this correlation, in this research, I designed and developed a comprehensive identification system exploring and including almost all applicable machine learning algorithms which work with high performance in this area of study and fit very well in the main, relevant parts of a thorough identification system, such as PRPD data pre-processing and feature generation, PRPD feature extraction, and PRPD pattern classification parts. This system aims at identifying multiple, simultaneously activated PD sources, as well as single PD sources. These multi-sources of PD sometimes occur in HV insulation systems and identification of them is very practical. To perform classification, a novel algorithm was developed in addition to the algorithms which were implemented for single PD source classification. This novel algorithm is powerful in classification of multi-source PDs, similar to the algorithms which are available in the identification system and operate considerably successful on single PD source classification. This novel algorithm was appended to the identification system to enhance its performance.

To test the performance of the system for identification of PD sources, different test setups to model various artificial single source and multi-source of PDs in different insulation media were designed. Obtaining accurate and reliable measurements and using them in the proposed system was one of the most important properties of this research. To start analyzing the measured data, PRPD patterns were recorded for single and multi-source of PDs, then different parts of classification system were optimized using the generated dataset as the input of the system. In general, the proposed classification system was designed to conduct prosperous automated identification of PD sources based on the availability of high-speed data processors and well-developed statistical techniques in machine learning. Availability of this reliable, automated, identification system for

single and multiple PD sources enables online condition assessment monitoring of high voltage (HV) apparatus more accurately and efficiently. In this way, the ability to identify defects in early stages is possible that would enhance safety of HV apparatus, such as transformers, electric machines, cables, and Gas-Insulated Switchgear (GIS).

1.3 Methodology and Contributions

This PhD thesis presents an automated classification system[2] for both single and multiple, simultaneously activated PD source identification that will investigate the relationship between the variation of PRPD patterns and the sources of PD. Such system consists of feature extraction methods and classifier algorithms that are implemented for the recognition of the source of partial discharge. Once the phase resolved PD patterns are recorded, features are generated. Due to the "curse of dimensionality" [2] and to increase the processing speed, and to reduce the required memory, different methods of dimensionality reduction to extract features that represent the fingerprints of the PD source are employed. This removes redundant and ineffective information and decreases the number of features while still capturing high portion of information [2, 3].

This PhD thesis contributes in the areas of:

- Classification of single PD sources using 12 high performance, applicable methods on PRPD pattern data for dimensionality reduction (including the traditional statistical operators) which are chosen exploring almost all available well-developed feature extraction techniques, as well as 10 well-known algorithms for classification have been explored. The classification success rate of their application on the PD patterns of the discharge activities in different insulation media including air, oil, SF_6 has been evaluated.

- Some of the classifier algorithms developed in this work, such as fuzzy classifiers, are not only capable to show high classification accuracy rate, but they also calculate the "degree of

[2]The codes have been developed in Matlab

membership" of a sample to a class of data. This enables probabilistic interpretation of a new PRPD pattern that is being classified.

- The availability of this degree of membership for future PRPD samples would allow safer decision making based on the risk associated with different sources of PD in HV apparatus.

- Test sets are designed to study PRPD patterns and show the performance of proposed classification system on identification of single partial discharge sources in oil-immersed insulation under electrical stress and the power transformer cellulose insulation samples under both electrical and thermal stresses. This capability enables online monitoring of high voltage cellulose insulation more accurately and efficiently which helps to prevent most transformer failures.

- To generate a dataset, two commonly-used feature generation approaches which have been used in the past [4] are modified in this work to considerably increase their discriminatory power.

- A new approach for feature generation is proposed with strong discrimination power to differentiate between PRPD patterns of different sources. The efficiency of this approach will be considerably perceived dealing with multiple simultaneously activated PD sources in HV insulation.

- Classification of the PRPD pattern that is a mix of multiple, simultaneous PD sources is performed. To do so, I developed a novel algorithm to identify multiple, simultaneously activated PD sources using PRPD patterns that are widely used in power industry and are easier to analyze compared to PD pulse waveforms analysis. The multi-source PRPD pattern classification is developed using training and test databases that are generated from fingerprints of single-source PD patterns and probabilistic interpretation is performed following a novel two-step Logistic Regression (LR) algorithm [6].

In summary, proposed classification system has several advantages in HV insulation monitoring

which are:

* Classification of PRPD pattern with high accuracy rate.
* Probabilistic interpretation based on the "Degree of Membership."
* Risk assessment of different sources of PD in HV apparatus.
* Investigation of similarity between different sources of PD.
* Referring of a marginal classification to an expert operator.
* High classification rate of multiple, simultaneous PD sources.
* Enabling continuous 24/7 monitoring.
* Considering the effects of different parameters such as the increase trend of work temperature on generated PRPD patterns.

The outcomes of this research have been published in three conference papers and three IEEE Transaction journal papers on DEIS and Power Delivery have been submitted [7–11].

Journals:

1. H. Janani; M. Jafari-Jozani; B. Kordi,"Classification of Simultaneous Multiple Partial Discharge Sources Based on Probabilistic Interpretation Using a Two step Logistic Regression Algorithm," *IEEE Transactions on Dielectric and Electrical Insulation*, Accepted for publication on July 2016.

Conferences:

1. Janani, H.; Jacob, N.D.; Kordi, B., "Partial Discharge Pattern Recognition for Thermally-degraded Cellulose-oil Insulation," *CIGRÉ Canada*, Winnipeg, Manitoba, August, 2015.

2. Janani, H.; Jacob, N.D.; Kordi, B., "Automated recognition of partial discharge in oil-immersed insulation," *2015 IEEE in Electrical Insulation Conference (EIC)*, vol., no., pp.467-470, 7-10 June 2015.

3. Janani, H.; Kordi, B., "Remote Detection and Statistical Classification of Partial Discharge," *Cage Club Student Conference (CCSC) on High Voltage Engineering and Applied Electrostatics*, University of Waterloo, Waterloo, Ontario, August, 15^{th} 2013.

Chapter 2

Background and Literature Review

2.1 Partial Discharge (PD)

According to the IEC60270 standard, "partial discharge is a localized electrical discharge that only partially bridges the insulation between conductors and which can or cannot occur adjacent to a conductor. Partial discharges are in general a consequence of local electrical stress concentrations in the insulation or on the surface of the insulation. Generally, such discharges appear as pulses having a duration of much less than 1 μs. More continuous forms can, however, occur, such as the so-called pulseless discharges in gaseous dielectrics. Partial discharges are often accompanied by emission of sound, light, heat, and chemical reactions." [12]. Partial discharges can lead to both physical and chemical deterioration of insulation materials and if are not detected for a long time, they might eventually cause electrical breakdown of the HV equipment. Partial discharges may occur in different forms of internal discharge, surface discharge, and corona. Internal discharge refers to the discharges in cavities within solid insulation or at the edges of conductors in solid or liquid insulation. Surface discharges are a type of discharge may occur on the surface of insulation material. Corona refers to partial discharge in gasses around conductors that are remote from solid or liquid insulation. According to [13], "Corona (in air) is a luminous discharge due to

ionization of the air surrounding a conductor caused by a voltage gradient exceeding a certain critical value". Partial discharges is a complex physical process with stochastic properties [14]. These discharges are accompanied by many phenomena such as, acoustic and electromagnetic waves, charge displacement, and light, which might be used for the detection of their sources. Therefore, techniques employed for PD detection are based on chemical, acoustic, optical, electrical, and UHF measurements. The electrical measurement of PD which is originated based on charge displacement in high voltage AC systems are widely used and is the focus of this PhD thesis. The measurement procedure and system calibration has been performed according to the IEC60270 standard by injecting a calibration pulse into the device under test. Before more details are discussed some important definitions are presented as follows.

- Apparent charge q: "Apparent charge q of a PD pulse is that charge which, if injected within a very short time between the terminals of the test object in a specified test circuit, would give the same reading on the measuring instrument as the PD current pulse itself. The apparent charge is usually expressed in pico-coulombs (pC). The apparent charge is not equal to the amount of charge locally involved at the site of the discharge, which cannot be measured directly" [12].

- Discharge inception voltage: "In practice, the inception voltage U_i is the lowest applied voltage at which the magnitude of a PD pulse quantity becomes equal to or exceeds a specified low value" [12].

- Discharge extinction voltage: "In practice, the extinction voltage U_e is the lowest applied voltage at which the magnitude of a chosen PD pulse quantity becomes equal to, or less than, a specified low value" [12].

2.2 PD Source Identification

Once partial discharges are detected and measured, the practical interest is to identify the source of the discharge and differentiate its pattern from that of any existing interference. In general, for identification of partial discharge, two different types of PD patterns, namely phase resolved and time resolved data set have been mainly analyzed [5]. In the phase resolved PD pattern, two important parameters both in reference to phase angles are discharge magnitude, and number count of discharges. Time resolved data sets consist of discharge magnitudes in reference to the time. To identify different sources of PD based on their patterns, signal processing and pattern recognition techniques are required to be employed. In the next two subsections, a review on application of some pattern recognition and signal processing techniques which have been used mostly in the last two couples of decades for identification of single and multi-sources of PDs is presented.

2.2.1 Single PD Source Identification

Partial discharge (PD) measurements were first carried out almost 80 years ago [15, 16], but were not considered seriously for the reliability assessment of HV insulations until 1980s-1990s [17]. In the last three decades, automated recognition of single source PD patterns has been progressively investigated and several signal processing methods and classification algorithms have been employed for the analysis of single source discharge patterns, such as, the relative identification factor [18], time-series analysis [19,20], artificial neural networks (ANN) [5,17,20–28], fuzzy algorithm [29,30], support vector machine (SVM) [31,32], hidden Markov models [33], statistical tools [4,34], inductive learning approach [35], Bayesian [31], and K means [36]. Okamoto and Tanaka were among the first experts who started to work in this area. With a simple computer-aided measurement system in that time, they tried to analyze statistical characteristics of PD pulses represented against phase-angle of the AC applied voltage. Based on this, they showed that some simple characteristics of $\phi - q$ distribution profile may be used to predict a treeing breakdown in solid insulation [37]. In [4], Krivda used Fisher Discriminant Analysis (FDA) and Principal Component Analysis (PCA)

for feature extraction. Application of wavelet transforms has also been shown to be useful for PD source recognition [25, 31, 38]. To improve the performance of ANN in the classification of discharge patterns, knowledge-based preprocessing method and time series approach have been presented [19, 21]. Classification of different sources of PD requires a database for training the classifier and testing. In almost all previous studies, such database is generated based on the measurements conducted on artificial defects that are implemented in controlled laboratory test cells [4, 17, 21, 22, 31].

2.2.2 Multiple Simultaneously Activated PD Source Identification

As already mentioned, in the last three decades, automated recognition of PD patterns has been progressively studied and several PD classification algorithms have been applied to analyze discharge patterns [4, 5, 17, 19, 21, 22, 24, 25, 27, 29–32, 34–37]. However, there are many practical situations where the interest lies in the identification of multiple, simultaneously activated PD sources in insulation [39]. Recently, identification of these types of defects is receiving more attention [39–43]. To enhance a PD classification system, these multi-source PDs need to be successfully classified. However, PRPD patterns associated with multiple simultaneously activated PD sources are often partially overlapped [39, 44] that makes them very hard to be appropriately classified using available methods in the literature. A few studies have been conducted in this regard, which are mainly based on analyzing the PD pulse waveforms in several exploration ways (signal processing techniques) attempting to separate individual concurrent pulse sources [9, 39–43]. Classification of these types of multi-source PDs is subsequently performed on each selected single-source PD using its corresponding PRPD sub-patterns. This is usually done under the assumption that there exists a relationship between the nature of PD sources and their generated pulse waveforms, which helps distinguishing different pulse waveforms originated from different sources.

2.3 Pattern Recognition–A Review

Pattern recognition is "the scientific discipline whose goal is the classification of objects into a number of categories or classes. Depending on the application, these objects can be images or signal waveforms or any type of measurements that need to be classified" [2]. A pattern classification algorithm consists of three main steps as shown in Figure 2.1 [2, 3]:

1. Data pre-processing
2. Feature extraction (dimensionality reduction)
3. Implementation of the classifier algorithm

Below is a description of each step.

2.3.1 Pre-Processing

Pre-processing stage includes feature generation which is important in any pattern recognition work. The goal of this stage is to discover informative representations of the obtained data which has been recorded. After that, processing should be performed on the features prior to their utilization. Such processing includes outliers removal, scaling of the features [2].

2.3.2 Feature Extraction

The first major problem in building a classifier is the curse of dimensionality [2] which should be resolved by selecting a good combination of available features by applying a dimensionality reduction method. Another reason for reducing the features is the need for less computational complexity, high speed of classification, and less required memory. A large number of features and a limited number of observations can also lead the learning algorithm to over-fit to noise [2]. In addition, more features will make training a classifier more difficult [2]. Moreover, implementation of feature extraction techniques leads to the removal of multi-collinearity which improves the performance of

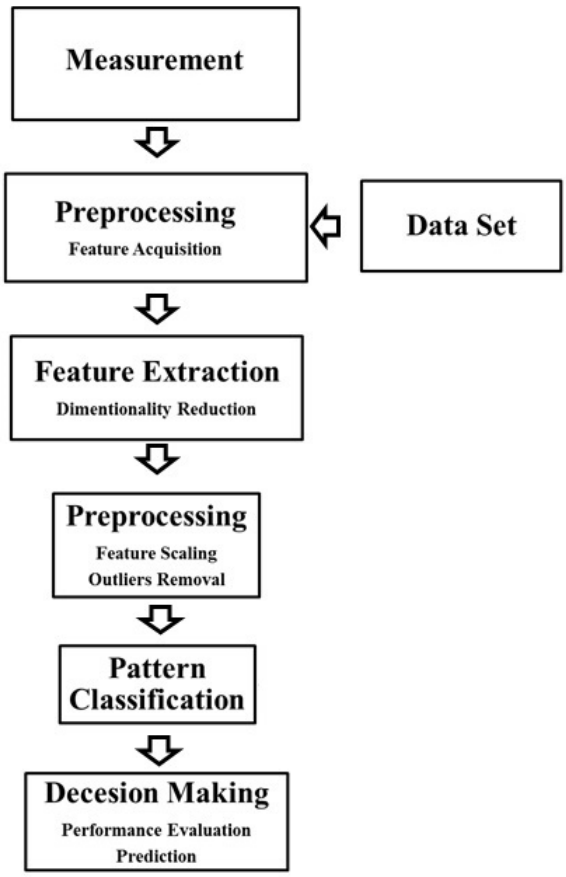

Fig. 2.1: General procedure of automated classification system.

the classification algorithm [3]. Multi-collinearity or collinearity is a phenomenon in which two or more variables in a regression model are highly correlated. To address these problems, one needs to select as many potentially-useful features as possible, and then reduce the number of features for classification. A limited number of dimensionality reduction techniques have been applied in classification of PDs in the past [4, 31], however during the last couple of decades, new linear and nonlinear algorithms for dimensionality reduction have been presented in the area of machine learning. These techniques attempt to extract and identify data resting on a low dimensional manifold of dimension K $(K < W)$, from a high dimensional space $\mathbb{R}^W$ that the manifold is embedded in

(W is equal to dimension of original space). "K" is typically referred to as the intrinsic dimension of the dataset [2].

2.3.3 Classifier Algorithms

Following the feature extraction step and construction of a set of training data from each of the PD sources, a classifier algorithm is required to find decision boundaries between classes in the low dimensional space. The classification stage comprises of performing two tasks; training (learning) and testing (classifying) [45]. Training task aims at partitioning the new low dimensional feature space, whereas the testing task is to assign the input pattern to one of the classes. Performance evaluation is then carried out based on the errors which might have happened in these assignments. The objective of designing a classification system is to predict and then assign future unknown samples that are probably different than the training data to one of the existing classes (and reject the marginal samples). The trained system should be efficiently optimized to show the desired performance in prediction of the test data. A highly-optimized classifier (to get maximum performance on the training dataset) sometimes results in undesired performance (overfitting) on the test set. Another problem that may occur during classification of test set is due to the large number of unknown parameters related to the classifier, such as the number of parameters in a large neural network [45]. Moreover, the ratio of the number of training samples to the number of features is an important factor. If this ratio is too small, it would influence the performance of the classifier (*i.e.* curse of dimensionality). A thorough investigation is required to design a powerful classifier for accurate PD source identification in different insulation media, using various algorithms for extracting features from PRPD patterns and building a number of classifiers.

2.4 Summary

In this chapter, a brief review of partial discharge definition and identification history of single source and multiple simultaneously activated PD sources in HV insulation was presented. Also

description of pattern recognition as the scientific discipline whose goal is the classification of objects into a number of categories or classes was explained. At the end of the chapter, three main steps of a pattern classification algorithm in addition to their role in the algorithm were briefly discussed. In the next chapter, application of pattern recognition with the implementation details of its different steps for identification of single PD sources in HV insulation based on using PRPD patterns is presented.

Chapter 3

PRPD Pattern Recognition

In AC systems, the phase resolved partial discharge (PRPD) pattern, which visualizes the occurrence of PD activities in reference to the phase of AC voltage, has been a valuable diagnosis tool (see Figure 3.1 for an example of a PRPD pattern). In a PRPD pattern, two important parameters, both in reference to phase angles, are the discharge magnitude and discharge rate. They form a bivariate distribution where each of the discharge magnitude and discharge rate can be separately analyzed with reference to each other and to the phase angle of the AC source. Below is a description of the pattern recognition steps presented in Chapter 2 when applied to a PD source classification problem using PRPD patterns.

3.1 PRPD Data Pre-Processing and Feature Generation

Measuring the PRPD patterns provides a bivariate distribution $H_{\mathrm{n}}(\varphi, q)$ that shows the correlation between discharge rate (n) , discharge magnitude (q) , and power frequency phase angle (φ) of the PD pulses. To generate a dataset from this bivariate distribution, the 2π phase angle window is divided into M phase windows and fingerprints are extracted from the PRPD pattern [4, 46]. Pre-processing stage includes feature generation which is important in any pattern recognition work. To generate one data point of the dataset, the PRPD pattern is recorded for T seconds

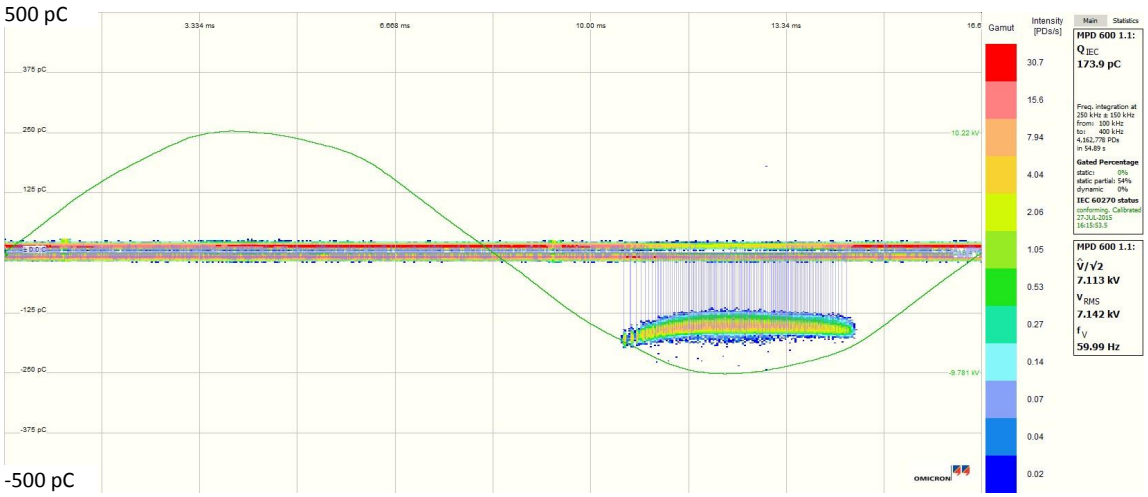

Fig. 3.1: PRPD sample for corona in air

and then the specific univariate distributions (will be explained in the following subsections) are evaluated. To generate a dataset of P points we have to repeat this process P times. Finally, the data points are formed in to a matrix whose dimension is $XM \times P$. In this work, typical values for these parameters are, $M = 100$, $P = 300$, X depends on the type of feature generation approach which is implemented (presented in follows), and $T = 3\,\mathrm{s}$ (or 180 cycles) for each type of the defects used for training and evaluating the classifiers. We normally allow 2 s between every two consequence data points. Optimum value for M is set equal to 100 based on a tradeoff between computational complexity and sufficient discriminatory information which will occur by lower and higher number of windows, respectively. After the generation of feature samples, some processing should be performed on those feature samples to make them ready for utilization. Such processing includes outliers removal and scaling of the features [2].

3.1.1 Available Feature Generation Approach

In each $2\pi/M$-wide phase window, parameters such as the average of discharge magnitudes, the maximum value of discharge magnitude, and the number of discharges are calculated. Considering these parameters in reference to the phase angle results in 3 univariate distributions of peak discharge $H_{\text{qmax}}(\varphi)$, average discharge $H_{\text{qmean}}(\varphi)$, and discharge rate $H_{\text{n}}(\varphi)$, respectively. As a result, X will be equal to 3.

3.1.2 Proposed Feature Generation Approach

In this thesis, a novel approach is proposed with strong discrimination power to differentiate between PRPD patterns of different sources. This approach is based on the application of q-quantiles on PD charge magnitudes of observations available in each of the M phase windows. Quantiles in PRPD pattern windows would be strong discriminators in separating different PD classes and help to better represent the shape of the PRPD pattern compare to the other two approaches (Results are shown in Chapter 5). This is because the number of PDs and their associated mean discharges are strongly influenced by the noises that are available in PRPD patterns. To make a powerful feature set, in addition to the application of quantiles, peak discharge, $H_{\text{qmax}}(\varphi)$ is also used. In this work, assuming $q = 200$, the PD magnitude observations in each phase window are divided into 201 groups with equal numbers of PD observations. Four highest quantiles (200^{th}, 199^{th}, 198^{th}, 197^{th}), $H_{200th-q}(\varphi)$, $H_{199th-q}(\varphi)$, $H_{198th-q}(\varphi)$, $H_{197th-q}(\varphi)$ plus pick discharge, $H_{\text{qmax}}(\varphi)$, are the five features generated from each phase window. In this approach, X will be equal to 5. The reason why this proposed method works with considerable high performance will be explained in the following sections.

3.2 PRPD Feature Extraction Techniques

For datasets like PD dataset with a large number of features and a limited number of observations, the feature extraction stage is very much needed [2, 3]. Having more information about a PRPD pattern seems to be useful, however, having many features compared to the number of observations is not efficient for producing a desired learning performance. Feature extraction techniques (interchangeably called dimensionality reduction) remove redundant and ineffective information and decrease the number of features while still the geometry of the data manifold is retained.

3.2.1 Dimension Reduction Algorithms

A feature extraction technique is a transformation method that transforms the data from the high-dimensional feature space to a new informative space with lower dimensionality [47]. In other words, such transformation transforms the matrix $X_{3M\times NP}$ to a matrix $Y_{K\times NP}$ where K is the number of features in the reduced (new) space, M is the number of windows in phase, and N is the number of classes. In this thesis, the implementation of 11 high performance dimensionality reduction techniques has been carried out. These techniques are applicable to PRPD datasets and are chosen exploring almost all available well-developed dimension reduction techniques as well as statistical operators. These techniques are divided into two main groups: 1) linear techniques that include Principal Component Analysis (PCA) [48], Fisher Discriminant Analysis (FDA) [49], and, 2) nonlinear techniques such as Kernel PCA [50], Kernel FDA (KFDA) [51], Metric Multidimensional Scaling (MDS) [52], Stochastic Proximity Embedding (SPE) [53], Isomap [54], Stochastic Neighbour Embedding (SNE) [55], and Local Linear Embedding (LLE) [56].

A third group can also be identified under the linear group that are linear algorithms derived based on the linear approximation of some of the nonlinear algorithms. These algorithms include Linearity Preserving Projection (LPP) [57], and Neighborhood Preserving Embedding (NPE) [58]. A summary of the dimensionality reduction techniques is shown in Figure 3.2. In general, time and memory which are required for execution of the linear algorithms are less than the nonlinear

algorithms, however execution time of all proposed algorithms could be considered in the same range (less than 1 minute in this research) except SPE and SNE which are iterative algorithms and they take longer time to generate lower dimensions of data (less than 5 minutes in this research).

3.2.1.1 Linear Techniques

Linear techniques for dimensionality reduction have been used in statistics for over a century. This type of techniques will map the data onto a low-dimensional feature space with a linear transformation which preserves some information of interest. Principal Component Analysis (PCA) [48], and Fisher Discriminant Analysis (FDA) [49] are the two famous algorithms of this group which are successfully implemented on PD type data in this research. A review of these two techniques has been presented in Appendix A.

3.2.1.2 Non-Linear Techniques

Non-linear techniques for dimensionality reduction have been developed in the last two decades. Data transformation procedure by these techniques is more complicated than the one with linear techniques. However, they work very well dealing with non-linear data manifolds. This type of techniques will map the data onto a low-dimensional feature space through different non-linear procedures based on their algorithms. This group of techniques is generally divided in two sub-groups, namely global group techniques which try to preserve global properties and local group techniques which try to preserve local properties (information) of data when they are mapped in the low dimensional feature space. Notably, local preserving techniques claim that by preserving properties of small neighborhoods around each data samples, the global properties of the data manifold will be also preserved. Non-linear group includes techniques such as Kernel PCA [50], Kernel FDA (KFDA) [51], Metric Multidimensional Scaling (MDS) [52], Stochastic Proximity Embedding (SPE) [53], Isomap [54], Stochastic Neighbour Embedding (SNE) [55], and Local Linear Embedding (LLE) [56] which are prosperously implemented on PD type data in this research. A

review of these techniques has been presented in Appendix A.

3.2.1.3 Linear Approximation Techniques (Modified Nonlinear Techniques)

This group can be identified under the linear group that are linear algorithms derived based on the linear approximation to some of the local nonlinear algorithms. These algorithms include Linearity Preserving Projection (LPP) [57], and Neighborhood Preserving Embedding (NPE) [58] which are also prosperously implemented on PD type data in this research. A review of these techniques has been presented in Appendix A.

3.2.2 Statistical Operators

One approach to generate features with discrimination power to differentiate between the discharge patterns of different PD sources is the use of several statistical parameters which can be applied on the univariate distributions. Both statistical operators that have been widely used in the literature (*e.g.* [5, 22]) for PRPD classification, and $q-$quantiles approach, which is introduced and applied in this thesis, to considerably increase the discriminatory power of statistical operators, have been used in this research. Some of these statistical operators, such as mean and variance, should be computed for both halves of the power cycle. Skewness and Kurtosis, on the other hand, are operators that should be computed with respect to a reference normal distribution [2]. One other feature is the number of local peaks in the univariate distributions in both positive and negative half cycles. Some operators have been used to evaluate the differences between the distributions in the half cycles of the power frequency. Discharge asymmetry, phase asymmetry, cross-correlation factor, and modified cross-correlation factor are in this group [2].

The q-quantiles have been used in this work as a novel approach with discriminatory capability to separate different PD classes. In this work, assuming $q = 3$, we divide the data set into four groups with equal numbers of data points (*i.e.* 0-25% , 25-50%, 50-75% and 75-100% of the total number of data points). Applying the 5 operators of mean, variance, skewness, Kurtosis, and the

number of peaks to both positive and negative cycles of $H_{\mathrm{qmax}}(\varphi)$,$H_{\mathrm{qmean}}(\varphi)$, and $H_{\mathrm{n}}(\varphi)$ results in 30 features for each PRPD pattern. Further, applying additional operators of discharge asymmetry, phase asymmetry, cross-correlation factor, and modified cross-correlation factors will generate extra 7 features [4]. In addition, application of 3-quantiles on both cycles leads to the generation of 18 more features. In total, a feature vector with $55 = (30+7+18)$ entries is constructed for each PRPD pattern. This vector can be used as the fingerprint of each discharge pattern for discrimination of different patterns.

The results of classifiers using statistical operators as selected features which has been mostly used in the past [4, 22, 31] will be compared with classifiers that use dimension reduction techniques. Comparison of the overall classification success rate related to the specific feature extraction/classification algorithms will help finding more efficient combination of algorithms in different insulation media.

3.3 PRPD Pattern Classification Algorithms

So far in this research, 10 well-known algorithms for the classification of PD sources have been studied. These algorithms are Support Vector Machine (SVM) [59], Kernel SVM [2], Fuzzy SVM [60], Fuzzy kNN [61], Multi-Layer Perceptron [62], Radial Bases Function Network [3, 62], Probabilistic Neural Networks [2, 63], Bayesian [2, 3], Naïve Bayes [2], and AdaBoost [64]. Some of these algorithms including fuzzy classifiers are not only capable to show high classification accuracy rate, but also calculate the degree of membership of a sample to a class of data beside assigning a class label. This enables probabilistic interpretation of a new PRPD pattern that is being classified. The availability of this degree of membership for future PRPD samples would allow safer decision making based on the risk associated with different sources of PD in HV apparatus. The fuzzy algorithms which have been used in this paper are Fuzzy kNN [61] and Fuzzy SVM [60]. A summary of the classifier algorithms is shown in Figure 3.2. The performance evaluation of all classifier algorithms integrated with different feature extraction algorithms on PD source identification is presented in

the following chapters. A review of classifiers which are implemented in this thesis and appended in the proposed classification system have been presented in the following subsections.

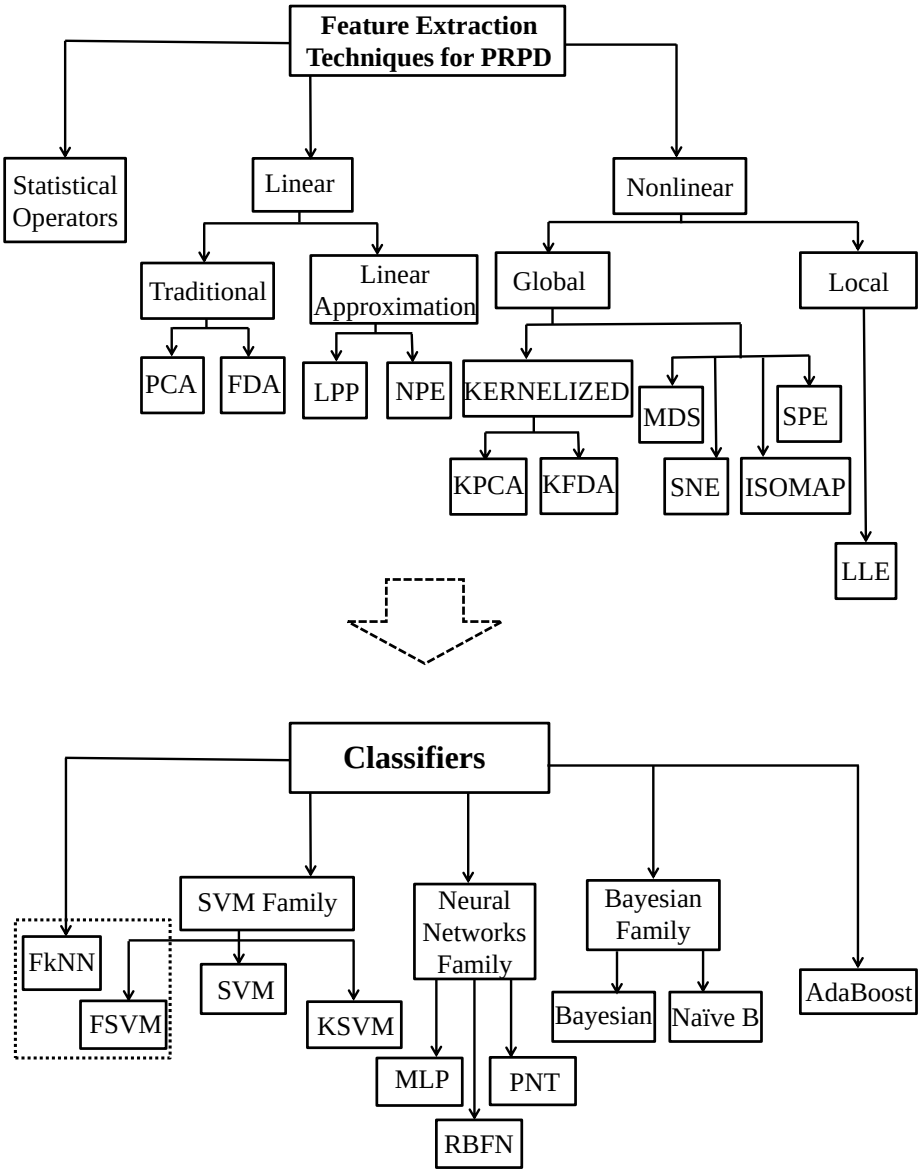

Fig. 3.2: PRPD dimensionality reduction and classifier algorithms.

3.3.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) [59] is a powerful algorithms for data classification. This algorithm aims at designing a hyperplane to classify different classes meanwhile maximizing the margin from classes. Such a hyperplane is more trustworthy when assigning new data to its belonged class. In general, SVM is an algorithm which tries to make the margin as large as possible and also tries to keep the number of points which are misclassified or located within the margin as small as possible. SVM algorithm is graphically described in Figure 3.3.

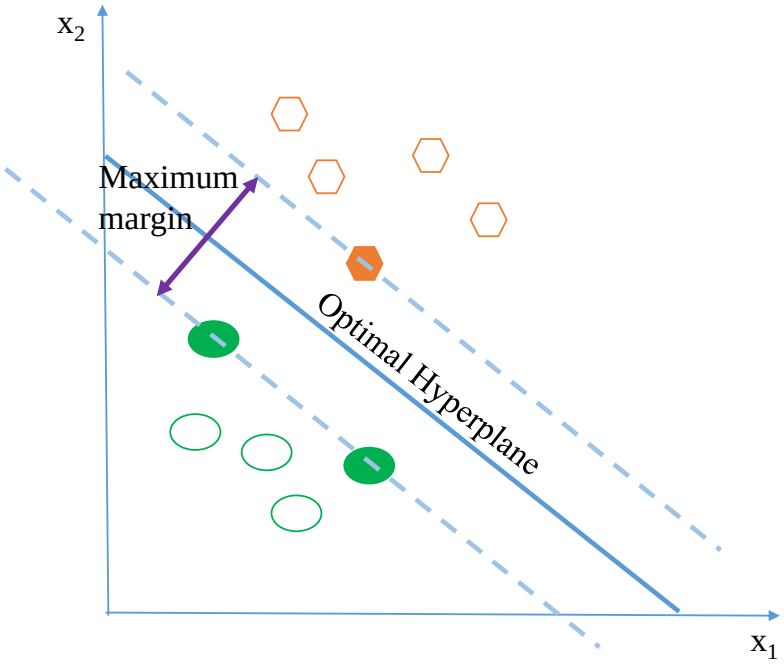

Fig. 3.3: Architecture of SVM algorithm.

3.3.2 Nonlinear Support Vector Machine

Nonlinear SVM is a kernelized version of SVM [2]. This algorithm starts with mapping the dataset (x_i) onto a higher dimensional feature space in which the classes can be classified by a hyperplane. KSVM begins by mapping the data with a feature map ϕ to a Hilbert space (H). However, if a kernel function $K(x, x_i)$ satisfies the Mercer's conditions [2], there would be a space in which $K(x, x_i)$ defines an inner product $< \phi(x), \phi(x_i) >$. Mercer's theorem doesn't mention how to construct ϕ (or even what H is), which is not actually an important issue. This is because K needs to be used in the nonlinear SVM algorithm, and we dont explicitly need ϕ. Finding an appropriate kernel implicitly defines a mapping onto a space with higher unknown dimensions. The curse of dimensionality in this way would be bypassed, as computational complexity does not depend on the dimensionality of the mapped space. This algorithm minimizes the cost function,

$$J(\omega, \omega_0, \xi) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{N} \xi_i \tag{3.1}$$

subject to

$$y_i[\omega^T \phi(x_i) + \omega_0] \geqslant 1 - \xi_i \quad and \quad \xi \geqslant 0, \quad i = 1, 2, ..., N,$$

where ω and ω_0 define the optimal hyperplane, in which they are a normal vector to the hyperplane and a bias, respectively. C is the regularization parameter, which decides the smoothness to balance misclassification and the margin maximization. The slack variables $\xi(\geqslant 0)$ are the error terms due to the misclassification. The above quadratic programming (QP) problem can be transformed into its dual form by the Wolfe dual optimization task, which results in

$$\underset{\lambda}{Maximize} \quad (\sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i,j} \lambda_i \lambda_j y_i y_j K(x_i, x_j)) \tag{3.2}$$

subject to

$$0 \leqslant \lambda_i \leqslant C, i = 1, 2, ..., N \quad with \quad \sum_{i=1}^{N} \lambda_i y_i = 0.$$

The Lagrange multipliers λ_i are related to the data points that are located within the margin or misclassified. Finally, the linear classifier in the mapped space assigns the class of each new data by calculating

$$g(x) = sgn[\omega^T \phi(x) + \omega_0] = sgn[\sum_{i=1}^{N} \lambda_i y_i K(x_i, x) + \omega_0)], \tag{3.3}$$

where $sgn()$ is the sign function. This classifier is a nonlinear one in the original space due to nonlinearity of the kernel function and so, it is suitable for application on overlapping clusters of data.

3.3.3 Fuzzy Support Vector Machine

Kernel SVM and in general SVM are powerful classifiers, however based on the formulations discussed above, each training point is considered belonging to one of the available classes and for each class of data, all data samples are treated uniformly. This is sometimes a drawback when dealing with noisy data, and when the effects of the training points are different. Many research have shown that the SVM is very sensitive to noises and outliers [60]. Also, sometimes, in a class of data, some training samples are more important than others and it should be considered in the classification. In that case, meaningful data samples must be classified correctly and not much care is required to be taken to some meaningless training samples such as, noises or outliers to whether or not classifying them correctly. The FSVM could be applied to reduce the effects of noises and outliers by setting the fuzzy membership to each data sample to emphasize the effects of more meaningful samples. As a simple example, this fuzzy membership could be a function of distance between the data samples and their corresponding class center.

The algorithm starts by applying a fuzzy membership to each data sample and reformulating kernel SVM into fuzzy SVM. In this way, different data samples can make different contributions in the learning of decision hyperplane. These fuzzy memberships are the weights w_is which are appended to each input data samples of x_i which belongs to the class of C_j. In general, for each

sample we would have $\{x_i, C_j, w_i\}$, where $0 < w_i \leq 1$ describe the degree of membership which the point x_i belongs to the class C_j. FSVM like KSVM, begins by mapping the data with a feature map ϕ to a Hilbert space (H) in which the classes can be classified by a hyperplane. To find an optimal hyperplane, this algorithm minimizes the cost function of,

$$\begin{aligned} & J(\omega, \omega_0, \xi) = (\frac{1}{2}\omega^T\omega + C\sum_{i=1}^{N} w_i\xi_i) \\ & subject\ to \quad y_i(\omega^T.\phi(x_i) + \omega_0) \geq 1 - \xi_i \quad i = 1, ..., N \\ & and \quad \xi_i \geq 0, \quad i = 1, ..., N, \end{aligned} \tag{3.4}$$

where ω and ω_0 define the optimal hyperplane, in which they are a normal vector to the hyperplane and a bias, respectively. Constant C is the regularization parameter, which decides the smoothness to balance misclassification and the margin maximization. Since the slack variables $\xi(\geq 0)$ are the error terms due to the misclassification, the term $w_i\xi_i$ is a measure of error with different weighting. The smaller w_i reduces the effect of the ξ_i such that the corresponding sample x_i would be treated as less important. The above quadratic programming (QP) problem can be transformed into its dual form by the Wolfe dual optimization task, which results in

$$\begin{aligned} & Maximize \quad (\sum_{i=1}^{N}\lambda_i - \frac{1}{2}\sum_{ij}\lambda_i\lambda_j y_i y_j K(x_i, x_j)) \\ & subject\ to \sum_{i=1}^{N}\lambda_i y_i = 0 \\ & and \quad 0 \leq \lambda_i \leq w_i C. \end{aligned} \tag{3.5}$$

The Lagrange multipliers λ_i are related to the data points that are located within the margin or misclassified. The sample x_i with the corresponding $\lambda_i > 0$ is called a supprort vector. There are two types of support vectors, one with corresponding $0 < \lambda_i < w_i C$ lies on the margin of the hyperplane and the other one with corresponding $\lambda = w_i C$ is misclassifed. The difference between

FSVM and KSVM (or SVM) is that the samples with the same value of Lagrange multiplier may indicate the different type of support vectors in FSVM due to the weight parameter w_i. Finally, the linear classifier in the mapped space assigns the class of each new data by calculating

$$g(x) = sgn[\omega^T \phi(x) + \omega_0] = sgn[\sum_{i=1}^{N} \lambda_i y_i K(x_i, x) + \omega_0)], \tag{3.6}$$

where $sgn()$ is the sign function. However, this classifier is a nonlinear one in the original space due to nonlinearity of the kernel function.

In this research, the sample weights are considered an exponentially decaying function [65] given by

$$w_i = \frac{2}{1 + exp(\beta d_i^s)}, \tag{3.7}$$

where $d_i^s = \|x_i - cc_j\|^{1/2}$ is the Euclidean distance between the data point x_i and its corresponding class center cc_j (geometrical mean of the data samples which belongs to the same class). β determines the steepness of decay. Kernel function which is used in this research is Gaussian kernel $K(x_i, x_j) = exp(-\|x_i - x_j\|^2 / 2\sigma^2)$, where σ is called the standard deviation.

3.3.4 k-Nearest Neighbor (kNN)

k-Nearest Neighbor (kNN) [2] is a non-parametric algorithm which is among the simplest classification algorithms. kNN classifies a new data point based on majority vote of the data points which are its neighbors. Various types of distances can be used for finding the k-nearest neighbors of a data point including Euclidean and Mahalanobis distances [2]. Finding the nearest neighbors among the entire dataset is an important problem regarding to this algorithm. In general, kNN shows better performance when the ratio of training samples compared to the dimensions of the feature space increases.

3.3.5 Fuzzy k-Nearest Neighbor (FkNN)

kNN classifier normally treats all the data samples equally important in the assignment of the class label to the unknown sample. This makes difficulty in classification when the classes of data overlap. In kNN, all samples are used in classification are given the same values of weights. Another drawback of kNN algorithm is its inability to estimate the degree of membership of an unknown sample to its assigned class. The fuzzy sets theory is introduced to the kNN technique to develop the fuzzy kNN algorithm and to solve the drawbacks which are associated with the kNN algorithm. This fuzzy kNN outperforms kNN not only by indication of lower error rate, but also the result of classier give a confidence measure in the classification of new sample by producing these membership values in addition to the label. This advantage could be emphasized in case the new data sample is not member of available classes. In general, fuzzy kNN implies to what degree the new sample belongs or does not belong to each class.

This algorithm starts by assigning class membership (weights) to all samples available in classes of data by different techniques. Then like kNN, this algorithm searches for the k labeled samples of nearest neighbors and assign the degree of membership for a new sample x belonging to all classes of data $u_i(x)$ by

$$u_i(x) = \frac{\sum_{j=1}^{k} u_{ij}\left(\frac{1}{\|x-x_j\|^{\frac{2}{(m-1)}}}\right)}{\sum_{j=1}^{k}\left(\frac{1}{\|x-x_j\|^{\frac{2}{(m-1)}}}\right)}, \tag{3.8}$$

$u_i(x)$ is the degree of membership of a new sample x to class i. The degree of membership of selected k nearest samples is shown by u_{ij} which is degree of membership of sample j belonging to class i. The variable m determines how heavily the distance is weighted when calculating each neighbors contribution to membership value.

3.3.6 Multi-Layer Perceptron

The Multi-Layer Perceptron [62] as displayed in Figure 3.4 is another algorithm which can be used for classification of discharge patterns. MLP is constructed of different layers, namely, input, hidden and output layers. The hidden layer and output layers consist of a number of neurons. In this network, equation can be written considering one layer of hidden layer, d input nodes, M hidden neurons, and C different classes of data as

$$a_j = \sum_{i=1}^{d} w_{ji}^{(1)} x_i + w_{j0}^{(1)} \tag{3.9}$$

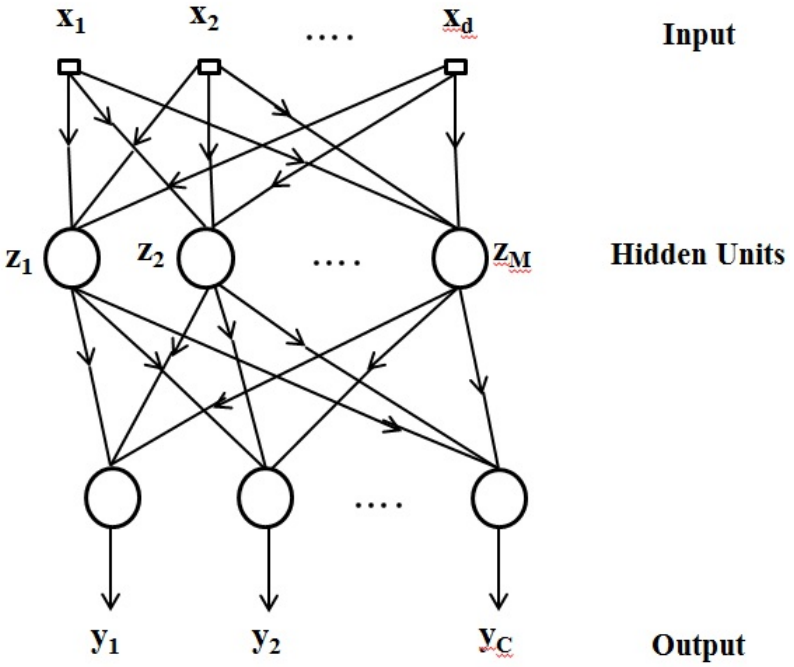

Fig. 3.4: Architecture of a multilayer Perceptron.

a_j is the output of the j_{th} hidden unit and it is formed by a weighted linear combination of the d input values which is added to a bias $w_{j0}^{(1)}$. $w_{ji}^{(1)}$ represents a weight in the first layer from input

i to hidden unit j. The activation of each output of the units is carried out by sigmoidal functions given by

$$Z_j = \frac{1}{1 + exp(-a_j).} \tag{3.10}$$

Subsequently, the output of the network is derived by a linear combination of the outputs of hidden units as

$$a_k = \sum_{j=1}^{M} w_{kj}^2 Z_j + w_{k0}^{(2).} \tag{3.11}$$

The activation of the k_{th} output neuron is performed by a softmax activation function to give

$$y_k = \frac{e^{(a_k)}}{\sum_{k=1}^{M} e^{(a_k)}.} \tag{3.12}$$

This algorithm also can be extended by considering a network with extra hidden layers; however the two layers perceptron can cover almost all types of recognition with good performance. The network used in this work is trained with scaled conjugate gradient backpropagation [62].

3.3.7 Radial Basis Function Network

In this algorithm, the output is found as the summation of weighted output of radial basis functions [3, 62]. In fact, each basis function works like a hidden neuron so the output is made from the combination of these basis functions as

$$y_k(x) = \sum_{j=1}^{M} w_{kj}\phi_j(x) + w_{k0} \tag{3.13}$$

ϕ_j is based on the Euclidean distance of the input vector x from a center which justifies the name radial basis function. The data in each class is modeled by a single kernel function as $p(x \mid C_k)$ which C_k is the label of different classes. The goal of classification is to find the posterior probabilities $P(C_k \mid x)$ for all classes. This algorithm is designed such that the outputs of the network represent

approximation to the posterior probabilities. To do that, a set of M basis functions is used

$$p(x \mid C_k) = \sum_{j=1}^{M} p(x \mid j) P(j \mid C_k) \tag{3.14}$$

Using Bayes formula, the result would be equal to posterior probability in which the normalized basis functions and the second layer weights are given by $\phi_j(x) = P(j \mid x)$ and $w_{kj} = P(C_k \mid j)$, respectively. In the training of this algorithm, first, basis functions should be selected by an unsupervised training technique using the training data. Subsequently, when the radial basis functions become fixed, the second layer weights are found by minimizing the sum of squares error function.

3.3.8 Probabilistic Neural Networks

Probabilistic Neural Network (PNN) is a form of neural network which can calculate nonlinear decision boundaries approaching the Bayes optimal decision surface [2, 66]. The structure of this algorithm is similar to back propagation; however a statistically-derived activation function is utilized rather than the sigmoid activation function. The PNN is trying to make an efficient structure for implementing Parzen estimate of an unknown pdf and Bayesian rule of the posteriori probabilities ($p(w_i)p(x \mid w_i)$) which should be the biggest for class w_i to be selected. To do so, all the feature vectors have been normalized to unit norm and then utilized in the network. The number of hidden layer nodes and the number of training data N are set to be equal. The training of this network is performed based on the usage of the training data. The inputs of activation function of the K_{th} neuron in the hidden layer is

$$\gamma_k = \sum_{j=1}^{l} x_{k,j} x_j = x_k^T x. \tag{3.15}$$

Using the activation function (the Gaussian kernel), the output of the K_{th} node would be

$$Z_k = exp(\frac{\gamma_k - 1}{\sigma^2}). \tag{3.16}$$

Subsequently, the number of output neurons (C) which are used is equal to the number of classes available in the training data. Each one of the output nodes acts like the linear combiners of the Z_ks which are connected to it. Therefore the output would be

$$y_c = \frac{P(w_c)}{N_c} \sum_{i=1}^{N_c} Z_i \tag{3.17}$$

where N_c is the number of training data in the C_{th} class. According to the Bayesian classification, the new data is assigned to the class with the highest output. It is shown that back propagation model is much slower than this algorithm [62].

3.3.9 Bayesian Classifier

Bayesian Classifier [2,3] is another algorithm which classifies data x based on using the C conditional probabilities (or posteriori probabilities) $P(w_i \mid x)$, where $w_1, w_2, ..., w_C$ are the C available classes. This algorithm tries to compute and then compare these conditional probabilities. The computation of posteriori probabilities can be carried out using the likelihood function of w_i with respect to x as $p(x \mid w_i)$. As $p(w_j, x) = p(x)P(w_j \mid x) = P(w_j)p(x \mid w_j)$, Bayes formula would be

$$P(w_j \mid x) = \frac{p(x \mid w_j)P(w_j)}{p(x)} \tag{3.18}$$

where $p(x)$ is the pdf of x and it is calculated by $p(x) = \sum_{i=1}^{C} p(x \mid w_i)P(w_i)$. Therefore, posteriori probabilities can be achieved by calculating the second side of (3.18). Comparing the posteriori probabilities of different classes, the biggest one would be selected as the class which x belongs to it.

3.3.10 Naïve Bayes

Another simple algorithm based on Bayes theorem is called Naïve Bayes [2]. The basics behind this algorithm is to assume that different features x_k, $k = 1, 2, ..., l$, are statistically independent.

So the probability of $p(x \mid w_i)$ can simply be calculated by

$$p(x \mid w_j) = \prod_{k=1}^{l} p(x_k \mid w_j), j = 1, 2, ..., C \tag{3.19}$$

Based on the training set, features should be described by their distributions and different parameters should be estimated for those distributions. In this thesis, Gaussian Naïve Bayes has been applied on the dataset. The classification of new data $x = [x_1, x_2, ..., x_l]^T$ to the class w_0 has been done

$$w_0 = arg \max_{w_j} p(w_j) \prod_{k=1}^{l} p(x_k \mid w_j), j = 1, 2, ..., C \tag{3.20}$$

This algorithm somehow is robust to the independence assumption of the features; however decision on different method of pdf estimation is based on the expected accuracy. Another simplification method based on Bayes theorem is used based on the assumption of equiprobable classes with normal probability density functions and with the same nondiagonal covariance matrix [2, 3]. The Bayes rule for this case is equivalent to minimizing the Mahalanobis distance.

3.3.11 AdaBoost

The AdaBoost (adaptive boosting) algorithm [64] is one of the boosting algorithms which resolved plenty of practical problems concerning the boosting algorithms. AdaBoost uses a weak learning algorithm in a number of steps to boost it into a strong algorithm with good performance in classification. This algorithm uses a distribution or set of weights over all the training samples. These weights would be updated in each step and increased for samples which are incorrectly classified. This causes the weak learner to concentrate on the hard training data points. After defining the base classifier (h_t) which returns a binary class label, its accuracy can be measured by the error and the weights in each step are updated such that the weight of the examples which are misclassified would be increased and the weight of classified data would be decreased. At the end

the new test point will be classified.

3.3.12 Multinomial Logistic Regression

Multinomial logistic regression (MLR) [6] is a linear multiclass probabilistic classification algorithm which is well-known in the field of Statistic. MLR trains a model to find posterior probabilities of available classes in the training set based on a linear function of all input variables. In this model, summation of these posterior probabilities for all classes would be equal to 1. In the training set, $X \in \mathbb{R}^k$ with N number of samples from unknown probability distributions belonging to different C classes, $P(C_d|x)$ is considered as the posterior probability that sample x belongs to class w_d by considering the posterior probability of class C as the base class (optional between classes). MLR model would be estimated in the form of $(C-1)$ linear discriminant functions or logit stochastic models as

$$log(\frac{P(w_i|x)}{P(w_C|x)}) = \boldsymbol{\beta_i h(x)} = \beta_{0i} + \beta_{1i}x_1 + ... + \beta_{ki}x_k, \tag{3.21}$$

for $i = 1, \ldots, C-1$, where $h(x) = [1, X]$, $X = (x_1, \ldots, x_k)$ and k is equal to the number of independent variables and C is equal to the number of classes. A single functional form of each individual class posterior probability could be calculated as

$$\begin{aligned} \underset{j=1,\ldots,C-1}{P(w_j|x)} &= \frac{exp(\beta_{0j} + \beta_{1j}x_1 + ... + \beta_{kj}x_k)}{1 + \sum_{i=1}^{C-1} exp(\beta_{0i} + \beta_{1i}x_1 + ... + \beta_{ki}x_k)}, \\ P(w_C|x) &= \frac{1}{1 + \sum_{i=1}^{C-1} exp(\beta_{0i} + \beta_{1i}x_1 + ... + \beta_{ki}x_k)}. \end{aligned} \tag{3.22}$$

The total number of $(C-1) \times (k+1)$ weight parameters $\boldsymbol{\beta} = [\beta_1; \ldots; \beta_{C-1}]$ with $\beta_j = [\beta_{0j}, \beta_{1j}, \ldots, \beta_{kj}]$ will be estimated by training set samples using the Maximum Likelihood Estimation (MLE) and Iteratively Re-weighted Least Squares (IRLS) methods [6]. The class label of a new sample x_{new} can be assigned by the Bayes classification rule given by

$$\hat{w}(x_{new}) = \underset{i\in\{1,2,...,C\}}{arg\,max}\ P(w_i|\mathbf{X} = x_{new}; \boldsymbol{\beta}). \tag{3.23}$$

3.3.13 Multi-Class Kernel Logistic Regression

Kernel logistic regression (KLR) [6] algorithm is a kernelized version of the linear logistic regression which performs nonlinear probabilistic classification. This algorithm, like KSVM, applies its linear model on the data which are mapped with a feature map ϕ in the higher dimensional feature space via the so called "kernel trick". Mapping is implicitly performed by specifying the inner product between pairs of data through a kernel function. Linear model in the mapped space would be a nonlinear one back into original space. This would be beneficial for classification of data in a training set containing classes which are linearly non-separable [6, 67, 68]. Logit models in this algorithm would be written as

$$f_j(x) = log(\frac{P(w_j|x)}{P(w_C|x)}) = \boldsymbol{\beta_j \phi(\mathbf{X})} \tag{3.24}$$
$$j=1,...,C-1$$

The weight parameters would be estimated using a convex optimization problem [69]. $\boldsymbol{\beta}$ could also be written as

$$\boldsymbol{\beta_j} = \sum_{i=1}^{N} \alpha_{ij} \boldsymbol{\phi(\mathbf{x_i})}. \tag{3.25}$$

Therefore, the final form of the model for KLR would be equal to

$$f_j(x) = log(\frac{P(w_j|x)}{P(w_C|x)}) = \sum_{i=1}^{N} \alpha_{ij} \boldsymbol{\phi^T(x)\phi(x_i)} = \alpha_{0j} + \sum_{i=1}^{N} \alpha_{ij} K(\boldsymbol{x}, x_i), \tag{3.26}$$
$$j=1,...,C-1$$

where using the kernel trick, the inner product has been substituted by a kernel function which satisfies Mercers conditions [6]. KLR like LR, generates the single functional form of each individual class posterior probability as follow

$$\begin{aligned} \underset{j=1,\ldots,C-1}{P(w_j|x)} &= \frac{exp(f_j(x))}{1+\sum_{j=1}^{C-1} exp(f_j(x))}, \\ P(w_C|x) &= \frac{1}{1+\sum_{j=1}^{C-1} exp(f_j(x))}. \end{aligned} \tag{3.27}$$

The class label of a new sample x_{new} can be assigned by the Bayes classification rule, which is given by

$$\hat{w}(x_{new}) = \underset{i\in\{1,2,\ldots,C\}}{arg\,max}\ P(w_i|\mathbf{X} = x_{new}; \boldsymbol{\alpha}). \tag{3.28}$$

3.4 Summary

In this chapter, the application of pattern recognition techniques for the identification of PD sources in HV insulation using PRPD patterns was elaborated. The implementation of each step of PRPD pattern recognition algorithm was explained and the applicable techniques which are used in this thesis to make a powerful PD source classification system were elaborated. These steps are namely, PRPD data pre-processing and feature generation, PRPD feature extraction techniques, and PRPD pattern classification algorithms.

The feature generation step in this research includes two approaches namely, 1) available feature generation approach, and 2) proposed feature generation approach. The proposed feature generation approach was developed and presented in this research for the first time. PRPD feature extraction step is divided in to two main groups with several sub-groups. These two main groups are divided into dimension reduction algorithms and statistical operators. Each one of these groups consists of different techniques which were explained in different subsections of this chapter. At the end, the final step in PRPD pattern recognition which is application of PRPD pattern classification algorithms was explained. Different classifier algorithms which are implementable on PRPD pattern data and used in this research were briefly elaborated. In the next chapter, the application

procedure and the performance evaluation of the designed PD source identification system using PRPD patterns of different single-sources of PD are presented. These sources includes different artificial PD sources designed and built to model PD sources in a variety of insulation media and under different working temperatures.

Chapter 4

Single PD Source Identification Using PRPD Pattern

To test the designed, automated classification system in single PD source identification, various artificial laboratory test setups are designed and built to model different PD sources in a variety of insulation media and conditions. The first test setup includes test cells which are designed to model common sources of PD in air, oil, and SF_6. These test cells include 3 sources of partial discharge in SF_6 (floating electrode, moving particle, and fixed protrusion), 2 sources of PD in transformer oil (free particle and needle electrode), and corona in air. The second test setup is designed to test the classification system in the identification of partial discharge sources in oil-immersed insulation. These sources of partial discharge are simulated to model common PD sources in HV transformer which includes bubble wraps to model small air bubbles[1], needle electrode in oil to simulate point-plane discharge, and floating metal particles (shavings) in oil. The third test setup is designed to show the performance of proposed classification system on the power transformer cellulose insulation samples under both electrical and thermal stresses. This capability enables online monitoring of high voltage cellulose insulation more accurately and efficiently which helps

[1] Application of bubble wrap makes us capable to have more control over this source of PD.

to prevent most transformer failures.

The results show that the proposed classification system is well able to successfully identify the sources of partial discharge in different insulation media under different working temperatures. Availability of this classification system enables continuous 24/7 monitoring of equipment and helps to identify PD sources in early stages which leads to safe operation of HV apparatus. Providing the probabilistic interpretation based on the risk associated with different PD sources, a marginal classified PRPD sample by the proposed classification system will be referred to an expert operator to do a visual inspection and make a proper decision; otherwise, decision will be made by the classification system.

4.1 Experimental Procedure for Pattern Recognition

Experimental Procedure for Pattern Recognition of Single PD Source in Different Insulation Media

Classification of different sources of PD requires a database for training the classifier and testing. Such database is generated based on the measurements conducted on artificial defects that are implemented in small laboratory test cells. The measurement procedure and system calibration have been performed according to the IEC60270 standard [12].

Figure 4.1 shows the experimental setup that consists of a high voltage transformer energizing the test cell, a coupling capacitor, the quadruple impedance, and commercial PD measuring equipment (Omicron MPD600). Voltage levels of 20% and 50% above the inception voltage have been applied to different test cells and PRPD patterns have been recorded for each test sample.

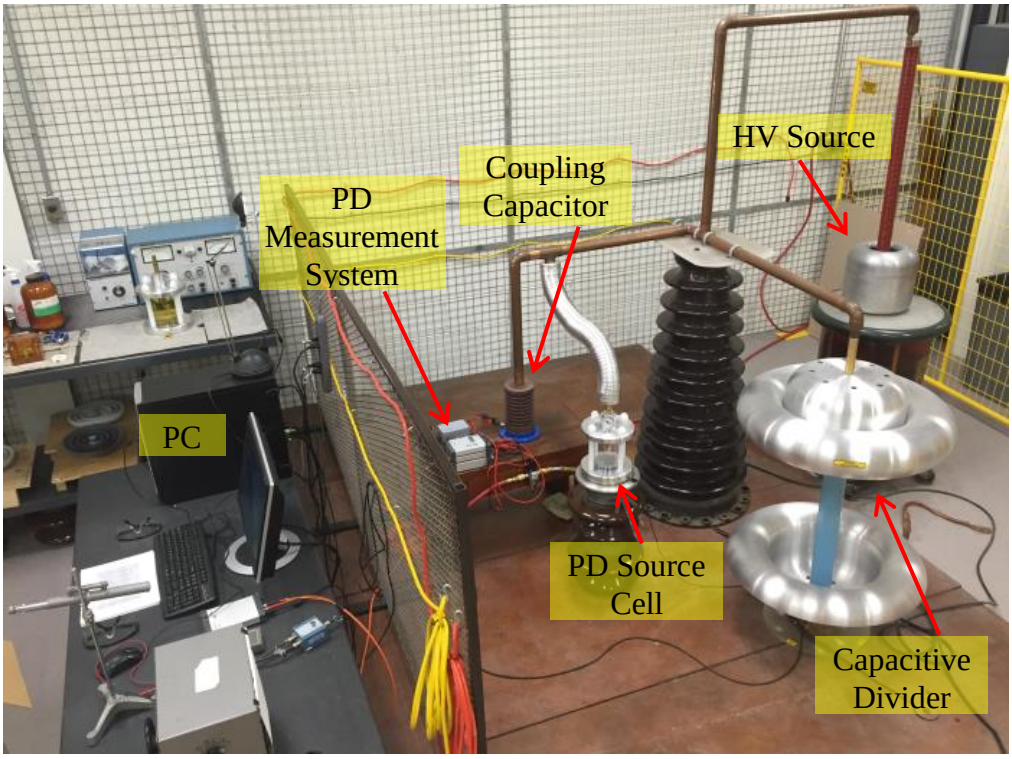

Fig. 4.1: Experimental setup of single-source PDs consists of an HV source, a coupling capacitor, a capacitive divider, a PD source cell, a PD measurement system, and a PC.

4.1.1 Test Cell Configurations

The first set of test cells designed in this research (see Fig. 4.2) was originally proposed by Hampton and Meats [70] and are shown in Figs. 4.3–4.5. These test cells are built to model different types of PD activities with different discharge mechanisms in air, oil, or SF_6. SF_6 test cells are designed to model common defects of GIS in small scale and be able to withstand a pressure of up to 500 kPa, consistent with gas pressure in gas-insulated switchgear (GIS). Sparking from a floating electrode, moving particles, and fixed protrusion are some of the major sources of PD in GIS [71] whose

laboratory models are shown in Figs. 4.3.a, 4.3.b, and 4.3.c, respectively. The test cells that can generate PD due to free particle in oil and needle electrode in oil are shown in Figs. 4.4.a, 4.4.b, respectively. Finally, Fig. 4.5 shows the test cell consisting of a needle and plane employed to generate PD due to corona in air at 100 kPa.

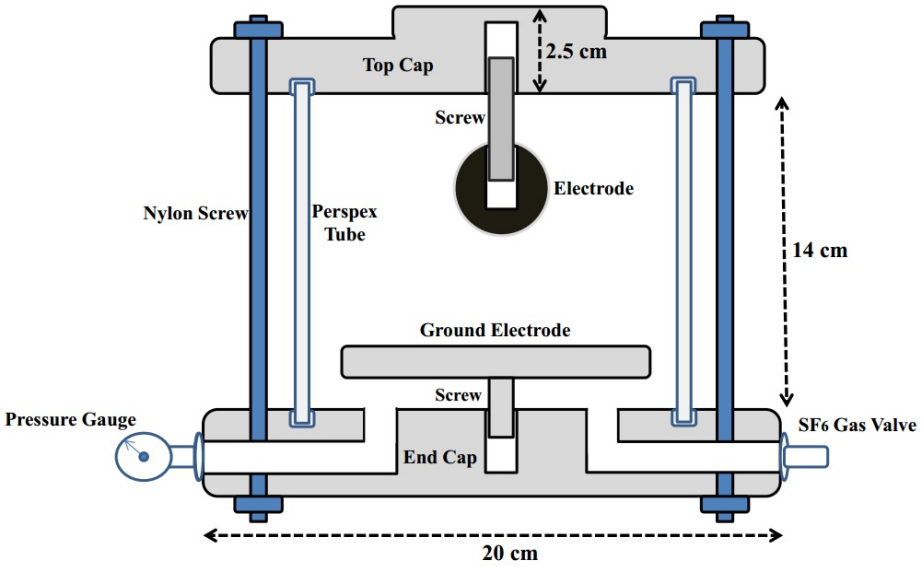

Fig. 4.2: Initial design of a test cell for GIS modeling, the perspex tube between aluminum top and end cap has been clamped by nylon screws prepare a pressurized vessel capable to withstand high pressure of up to 500 kPa.

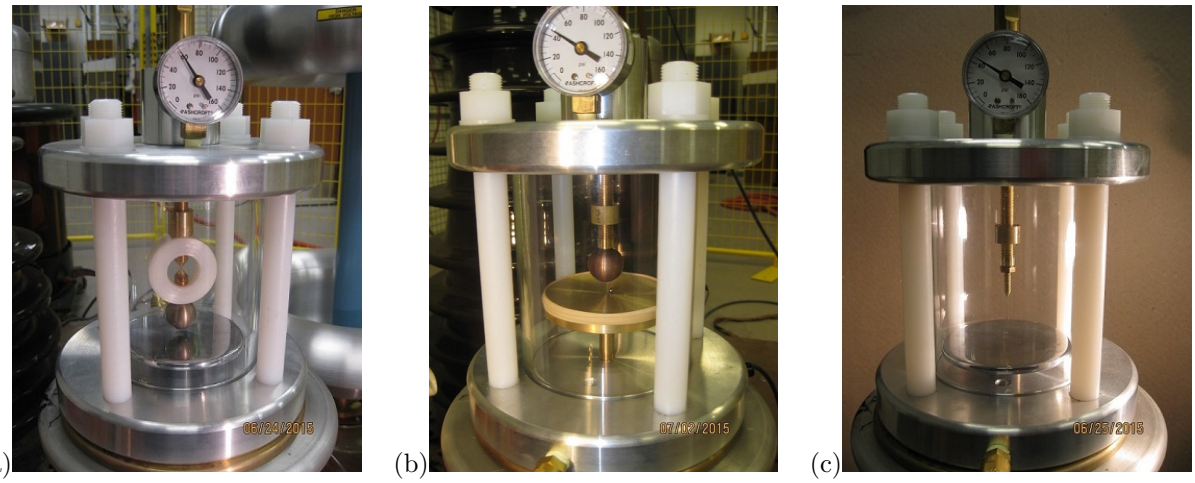

Fig. 4.3: SF_6 test cells; (a) floating electrode; (b) free particle; (c) point-plane electrodes. Each cell consists of a Perspex tube clamped by nylon screws between top and bottom aluminum caps that can withstand a pressure of up to 500 kPa.

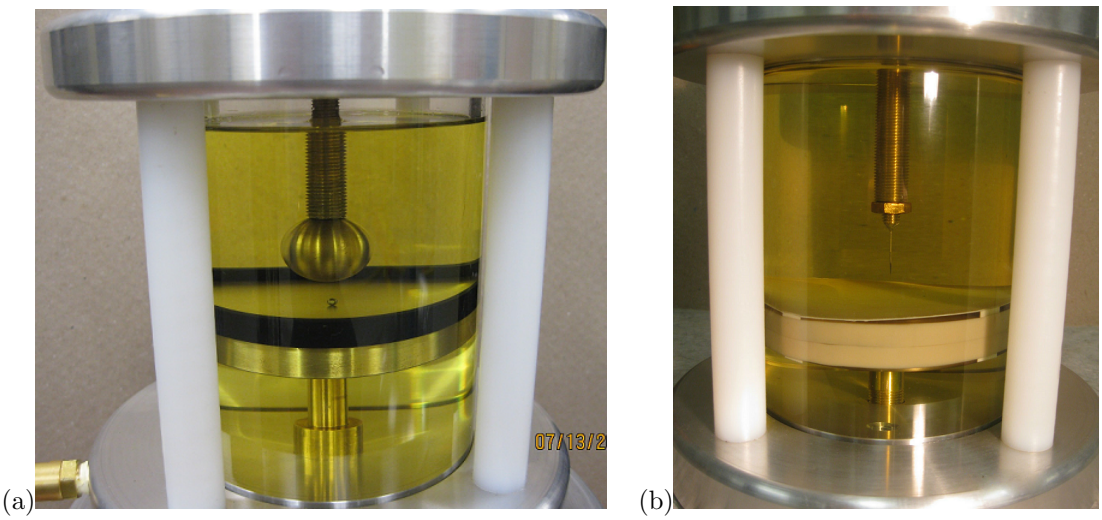

Fig. 4.4: Oil test cells; (a) free particle; (b) point-plane electrodes where the tip of the needle is $20\ \mu$m in diameter and the ground plane is covered with insulation paper to avoid breakdown.

Fig. 4.5: Point-plane electrodes in air.

4.1.2 Finite Element Simulation of the Electric Field and Voltage in Test Cells

One parameter which plays an important role in start and continuation of PD, is electric field and its configuration around the energized electrodes and ground. A commercial finite element software[2] is used to simulate the electric field and potential neighboring the electrodes in the cells. In the test cell for moving particle, high strength electric field is generated between the HV spherical electrode and grounded concave dish as shown in Fig. 4.6. A small 3 mm diameter aluminum bearing is located on the grounded dish. Once the test cell is energized, the small bearing starts moving and swinging on the dish. This movement leads to the generation of PDs. Basically, a free particle located in an electrostatic field stores electric charges and due to the Coulombs law, this electrically charged particle starts to move. This movement leads to discharge energy when particle contacts with the concave dish. These discharges would be the source of PDs. The floating electrode which

[2]Comsol Multiphysics

is not connected to HV electrode is going to be the source of corona. The simulated electric field distribution around the electrodes of this cell is shown in Fig. 4.7. Spacer inserts and corona shields are two the most common floating components in GIS [71–73]. To model another source of PD in GIS, a point-plane PD cell can be used to represent a fixed protrusion. Tip radius of the HV needle in this cell is about 10 μm which leads to a high electric field around it as depicted in Fig. 4.8. Other group of PD cells covers the mechanism of discharge in oil insulation materials and air. However, as electric field and voltage distributions are independent of the insulation material around the electrodes, these distributions for other insulation media of oil and air are only dependent on the electrode configurations and are similar to those of the same cell configurations in gas.

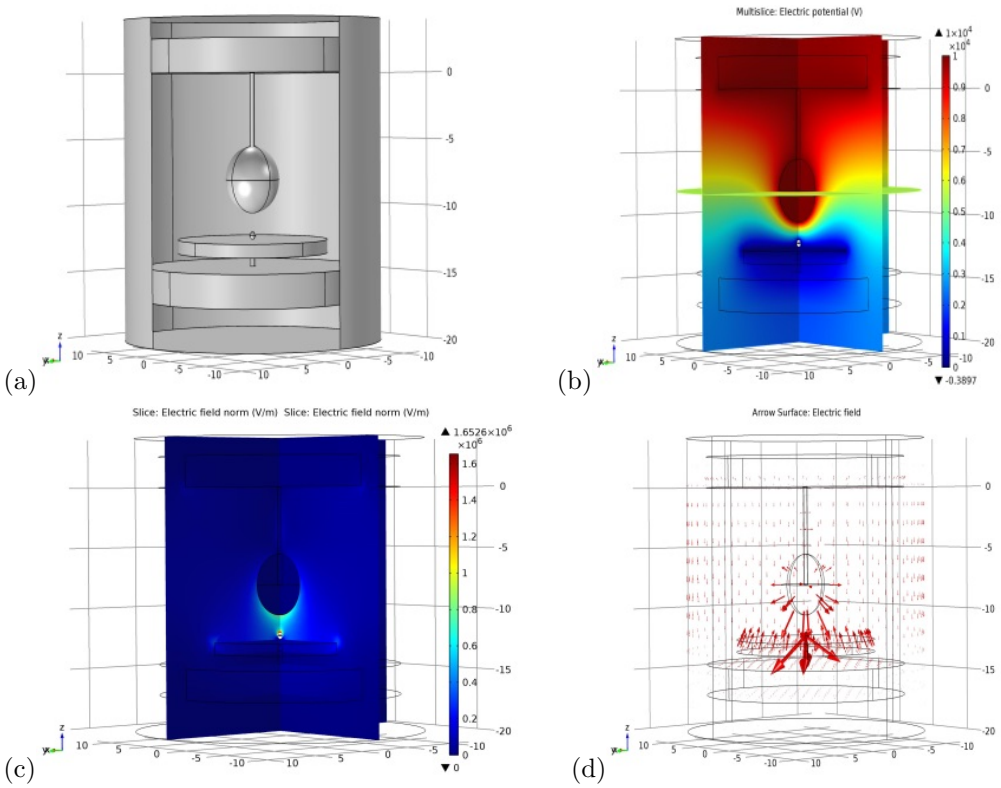

Fig. 4.6: Moving particle in SF_6 (a) General geometry in COMSOL, (b) Electric Potential (V), (c) Electric field norm (V/m), and (d) Arrow surface: Electric field

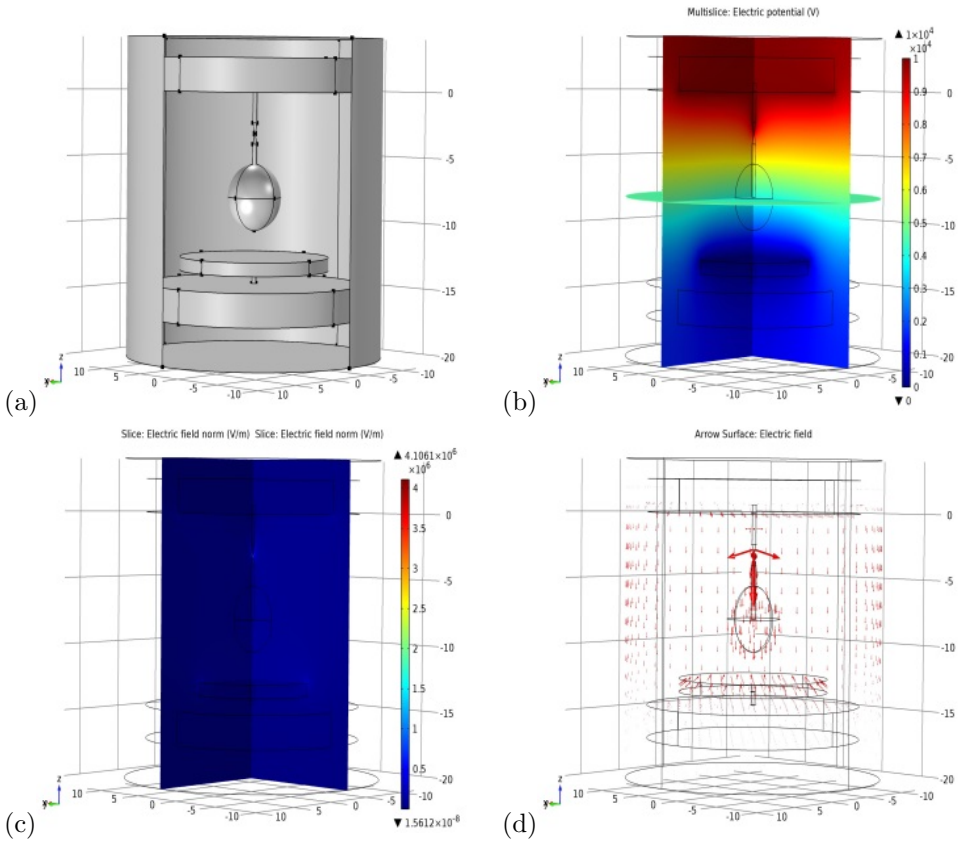

Fig. 4.7: Floating electrode in SF_6 (a) General geometry in COMSOL, (b) Electric Potential (V), (c) Electric field norm (V/m), and (d) Arrow surface: Electric field

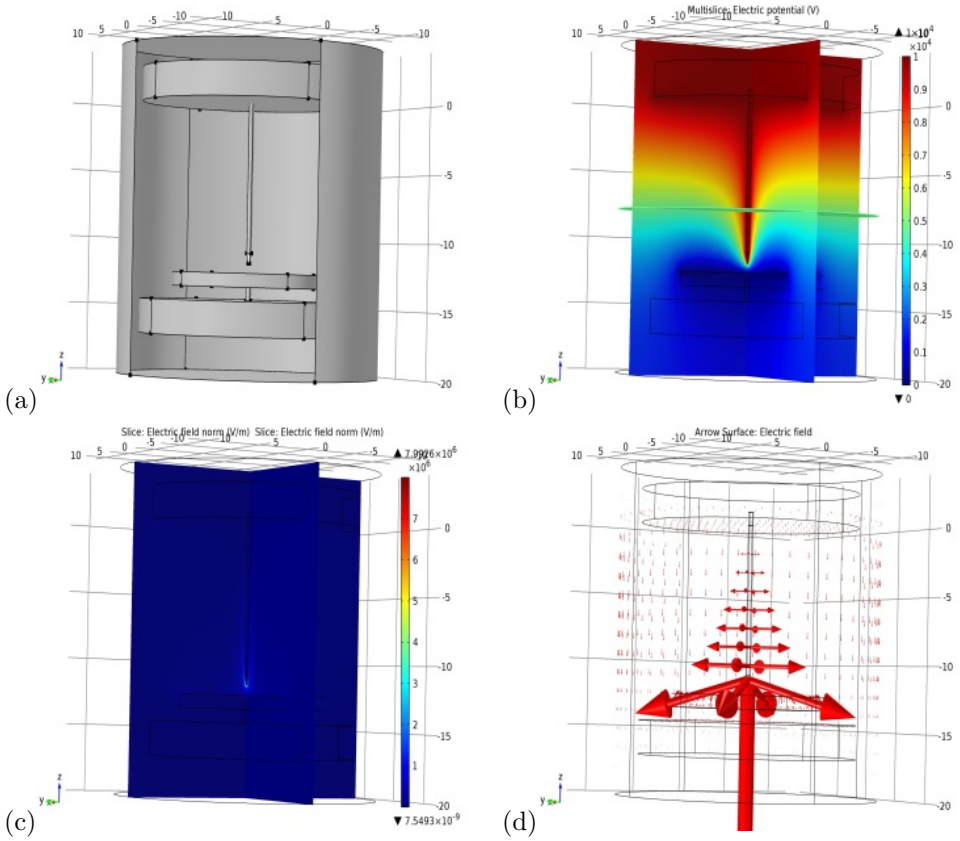

Fig. 4.8: Point-plane electrodes in SF_6 (a) General geometry in COMSOL, (b) Electric Potential (V), (c) Electric field norm (V/m), and (d) Arrow surface: Electric field

4.1.3 Results and Discussions

4.1.3.1 PRPD Patterns of Test Cells

The PRPD patterns of the six PD test cells are shown in Fig. 4.9 (3 dimensional PRPD patterns are shown in Fig. 4.10). For the floating electrode in SF_6 (see Figs. 4.3.a and 4.9.a), an inception voltage of approximately 15 kV was measured at 400 kPa. It was observed that the inception voltage and discharge magnitudes both increased with an increase in SF_6 pressure. It was also observed that both the inception voltage and PD magnitudes are strongly related to the gap size

between the energized electrode and the floating electrode, but not much sensitivity to the distance between the sphere and the ground electrode (5 mm in this experiment) was noticed.

The PRPD pattern shown in Fig. 4.9.b is related to a free particle in SF_6 at 400 kPa. This setup includes a small bearing with a diameter of 3.17 mm located on a concave dish ground electrode. The HV electrode is a 25.4 mm diameter sphere fixed at $\sim$ 10 mm from the ground electrode. As the voltage is increased to 10.5 kV, the small bearing starts to move across the plate toward the edge and back. This movement generates PDs between the bearing and the ground dish. PDs occur because of the charges that are transferred from the bearing to the ground electrode [71]. This experiment was repeated for different sizes of the bearing. When the size of bearing increases, the inception voltage decreases and the PD magnitude increases. However, if the size of the bearing increases to almost half of the gap distance, the movement will be a mix of swinging and bouncing when the bearing reaches the point right under the HV electrode. This is observable in the PRPD pattern too.

The PRPD pattern of a point-plane electrode in SF_6 at 300 kPa is shown in Fig. 4.9.c. To generate PD in this setup a tungsten needle with a tip radius of 10 μm located at a distance of the 15 mm from the ground plate has been used. This corona pattern is in fact the PRPD pattern of the namely positive corona in SF_6, *i.e.* it happens as the applied voltage increases somewhat higher than the negative corona inception voltage in SF_6 [74]. The typical PD magnitude of negative corona in SF_6 is in the range of -3 pC to -1 pC which happens in the negative half cycle of the applied voltage. Because of the low level of discharge of negative corona, in this work, we have considered positive corona only. Positive corona inception voltage was measured to be approximately 15 kV for this setup. However, discharge magnitude remains almost the same in the range between the inception voltage and approximately twice the inception voltage. Once the applied voltage is more than twice the inception voltage, the PRPD pattern and PD magnitudes start to show changes. There is no significant variation of PRPD pattern features with a variation in the SF_6 pressure.

The PRPD pattern of a free particle in oil is shown in Fig. 4.9.d. This setup and the electrodes

and the distances are the same as the setup of Fig. 4.3.b (except for the diameter of the bearing that is 2.77 mm. At a voltage of about 12.5 kV, the bearing is held right under the HV electrode with (almost) visible PD activities between bearing and the ground plane. The PD leads to the release of gas bubbles which move from the PD location toward the HV electrode. Sometimes the bearing starts to bounce for a short period of time that is visible to naked eyes. PD magnitude will increase as the size of bearing increases. In addition to the electric field enhancement due to the spacing changes, larger bearing can also store and transfer more charge to the ground electrode so PD magnitude becomes higher. Bouncing of a free particle under HV electrode will also get to be more pronounced for bigger bearings. Comparing this pattern to the same source of PD in SF_6, the spread of discharge in SF_6 can be explained by the movement of bearing on the ground electrode surface which leads to a bigger volume of discharge region.

To model point-plane discharges in oil, as another source of PD in oil, a 10 μm tungsten needle electrode configuration is used (Fig. 4.4.b). The high voltage is connected to the needle with its tip located 10 mm away from the grounded electrode. To avoid breakdown, grounded electrode is covered with a piece of insulating paper. The PRPD pattern of the needle electrode is shown in Fig. 4.9.e. Inception voltage of this test cell was 20 kV. It is observed that PD effects in this pattern are more vigorous in the positive half cycle with a large dispersion. The last PRPD pattern shown in Fig. 4.9.f is related to the corona in air. This setup is similar to that used for the generation of corona in SF_6 but this experiment is done in air at 100 kPa. The inception voltage of this test cell was 6 kV.

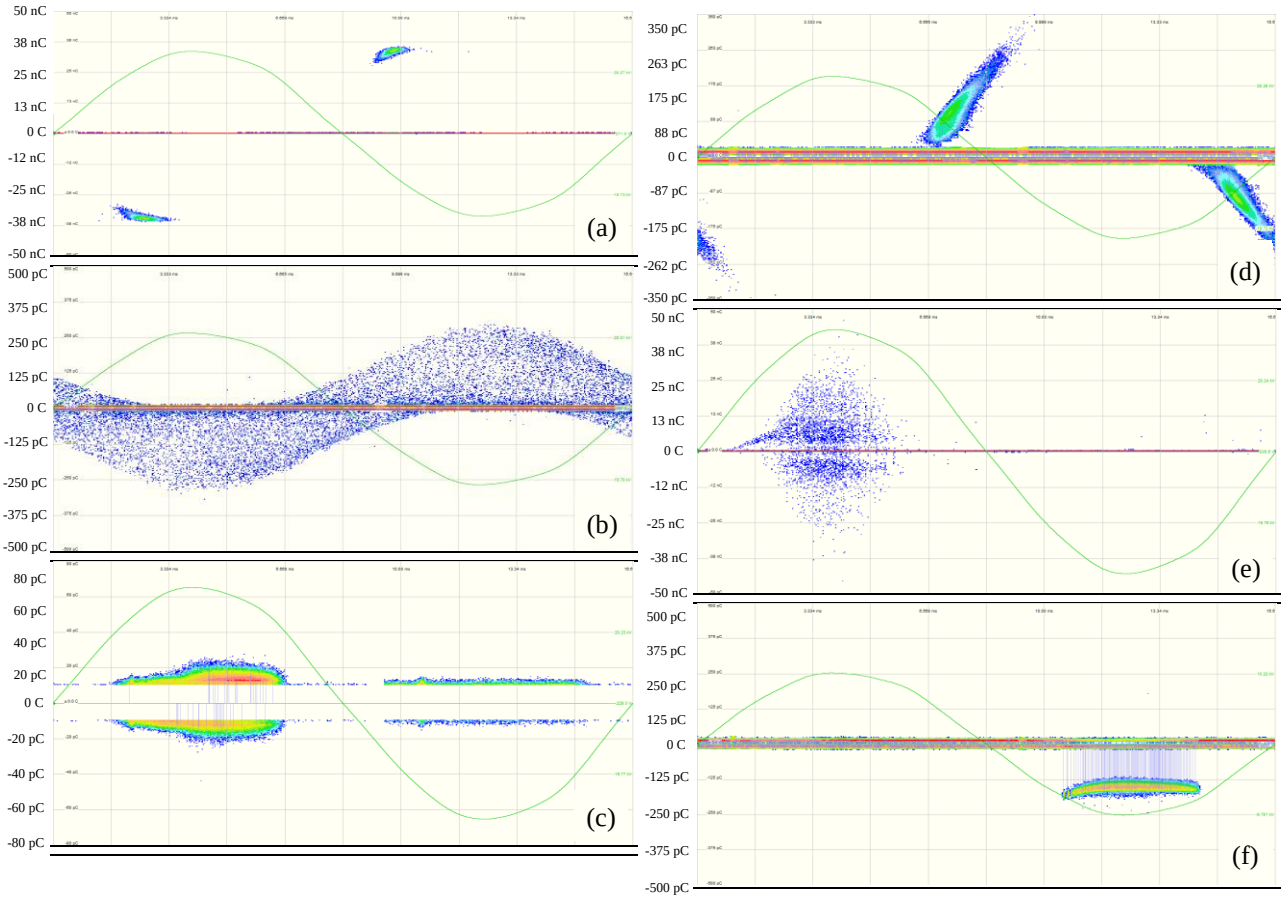

Fig. 4.9: PRPD patterns of, (a) floating electrode in SF_6; (b) free particle in SF_6; (c) point-plane electrodes in SF_6; (d) free particle in oil; (e) point-plane electrodes in oil; (f) point-plane electrodes in air.

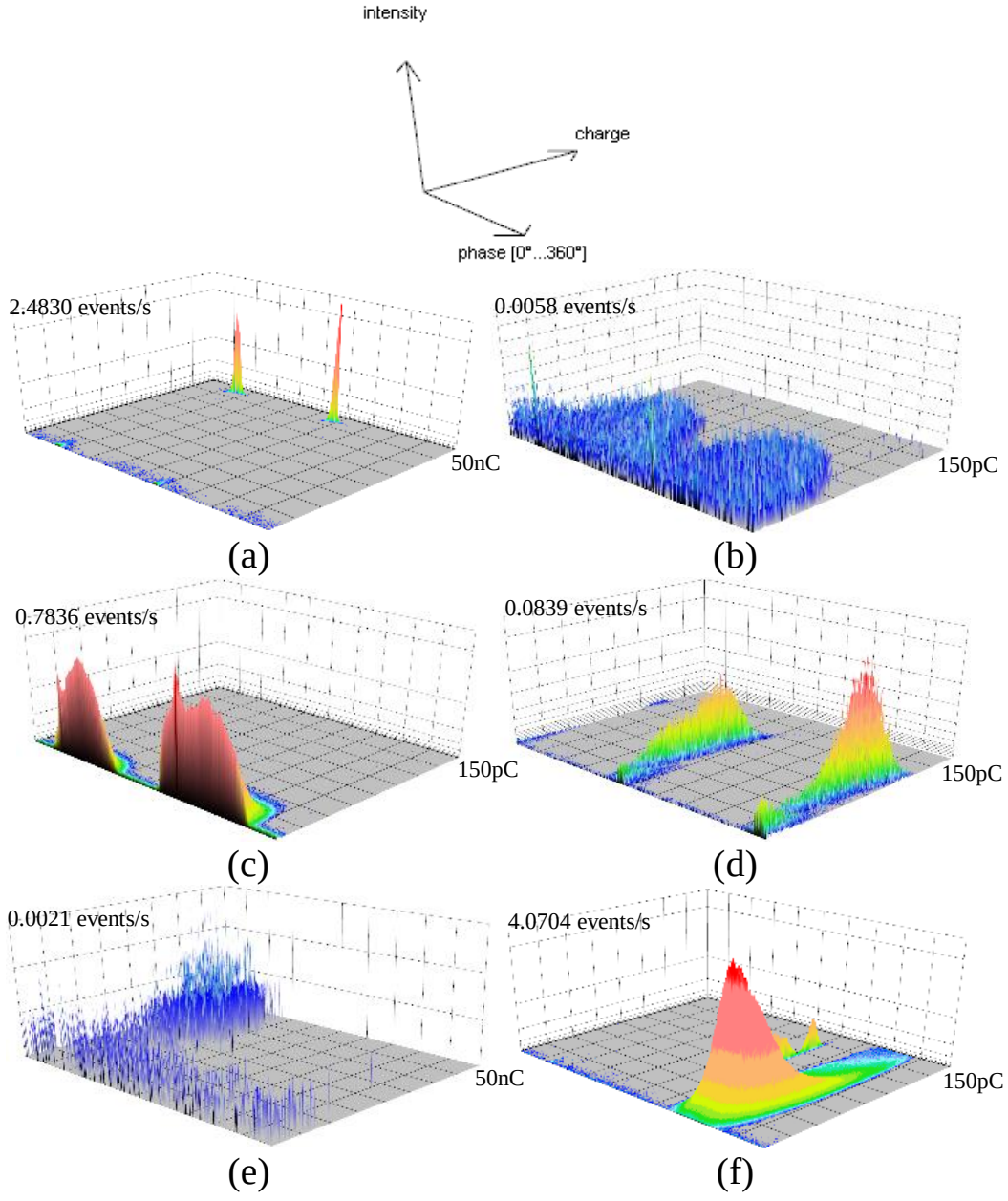

Fig. 4.10: 3 dimensional PRPD patterns of, (a) floating electrode in SF_6; (b) free particle in SF_6; (c) point-plane electrodes in SF_6; (d) free particle in oil; (e) point-plane electrodes in oil; (f) point-plane electrodes in air.

4.1.3.2 Classification Procedure

Using the experimental setups, a total of 300 data points are generated for each of the 6 different classes of PD sources [3]. For each type of the defects, the data points form the dataset matrix $\mathbf{X}_{3M\times NP}$ whose dimension is 300×1800 (*i.e.* $M = 100$, $P = 300$, $N = 6$). Application of the dimension reduction algorithms listed in section. 3.2.1 (except for FDA and kernel FDA) on matrix $\mathbf{X}$ results in a dimension reduction from 300 in the original space to 9 in the new informative space, *i.e.* $K = 9$. The new dimension $K = 9$ is the appropriate dimensionality of the reduced feature space that corresponds to the intrinsic dimensionality of the data determined by maximum likelihood estimation [2]. However, for FDA and kernel FDA, the new dimension is equal to $K = 5$. This number is selected based on FDA and kernel FDA algorithms which require the dimension be (at most) one less than the number of classes. In summary, the dimension reduction techniques (other than FDA and kernel FDA), FDA (including kernel FDA), and statistical operators reduce the dimension of the original datasets from 300 to 9, 5, and $55 = (30 + 7 + 18)$, respectively.

To perform PD source classification, the dimension-reduced dataset $\mathbf{Y}$ is fed to the classifier algorithms for both training and testing purposes. To do performance evaluation of each classifier algorithm, the classification error rate needs to be calculated. The classifier at first should be trained using training samples. Then, it has to be evaluated based on its classification performance on the test samples. The percentage of misclassified test samples is considered as an estimate of the error rate. To do so, and to also optimize the different classifier parameters, first the data in $\mathbf{Y}$ is split into two subsets; 80% for training and 20% for testing. The 80/20 ratio for testing and training is selected as a tradeoff; if the training set becomes small, classifier will not be very robust and if the test subset becomes small then the confidence in the estimated error rate will be low [45]. To run the optimization procedure for different classifier parameters, a $10-$fold cross validation (Rotation Method) is applied on the 80% training set [45]. An $n-$fold cross validation algorithm has been selected over leave-one-out or holdout methods because of its higher efficiency and better

[3] The partial discharge data recorded for 3 seconds (180 cycles of the applied voltage) generate one data point.

performance on the PD subset [45]. This method divides the training set into n subsets of equal size, uses $n-1$ subsets for training and one for testing. This procedure is repeated n times until all the training samples have been used for the training and exactly once for testing. In this work, this cross validation process has been repeated 10 times and 10 error rates have been averaged to produce a single classification accuracy rate of the algorithms on the training set. The optimal values for different parameters of classifiers will be found based on this classification accuracy.

After optimization, the classifier is trained using the whole training subset with the optimal value of parameters. At the end, to measure the performance evaluation of each classifier algorithm, the classification error rate is calculated by assigning a class label to the testing samples (*i.e.* the 20% that did not contribute in training and optimization process). To calculate a more accurate error rate for each classifier, the training/testing procedure of data splitting, cross validation, and testing has been repeated 5 times. Classification accuracy is averaged over the 5 trials and represents the success rate of each feature extraction/classifier algorithm.

Data samples for the 6 different classes of PD sources are captured under 2 levels of voltage equal to 20% and 50% higher than the inception voltage [4] [4] and two different noise levels[5]. The classification accuracy rate evaluated for the combination of each classifier integrated with different feature extraction algorithms are demonstrated in Tables B.1– B.12 in the Appendix . Each table presents the classification success rate related to the specific pair of feature extraction/classification algorithms for each individual source of PD. The last column of each table shows the overall classification success rate.

4.1.3.3 Performance Analysis of Classifiers

The results show that not only the nonlinear feature extraction algorithms work properly when applied on PD datasets, but also some of them outperform the classification results by linear

[4]These two levels of voltage are applied to make sure about the reliability of insulating materials for long time application and to analyze insulation performance during over-voltages which might happen due to Ferranti effect [75], short circuits and switching transients.

[5]Two noise levels are 4 pC and 8 pC.

algorithms and statistical operators. This advantage is because the nonlinear feature extraction algorithms are capable of dealing with complex nonlinear data manifold and work better with higher discrimination that leads to better performance of the classifier. Since different data samples from different sources are somehow mixed up with each other, a better performance from nonlinear algorithms is expected.

As it is displayed in Tables B.1– B.12, almost all algorithms result in a desirable classification accuracy, however FSVM, KSVM and AdaBoost classifier algorithms integrated with most of the feature extraction algorithms outperform the other 7 algorithms. Among these classifiers, Naïve Bayes shows less accuracy that is due to the basic assumption in this algorithm which is to assume that different features are statistically independent [2]. Also SVM attains lower classification accuracy compared to KSVM and FSVM because it is a linear classification algorithm and may not be able to deal with the nonlinearity of the data samples. Despite its simple architecture, kNN shows a good performance along with different feature extraction algorithms. These tables show that classifier algorithms work with high accuracy when integrated with MDS, KPCA, Isomap, SNE, and LPP.

The results also show FSVM and KSVM integrated with MDS outperform other feature extraction-classification algorithms with a classification rates of 99.4% and 99.1%, respectively. FSVM and KSVM start with mapping the dataset onto a higher dimensional feature space where in that space the classes can be classified by a hyperplane [2, 60]. The advantage of FSVM over KSVM is that the importance of some training points can be considered in the training process. This leads to make the classifier less sensitive to the effects of noise and outliers [60]. Classifier algorithms integrated with PCA and FDA from the linear group work with higher classification accuracy compared to statistical operators. However, classification using SPE as the feature extraction method does not show any desirable accuracy compared to other feature extraction algorithms.

4.1.3.4 Probabilistic Classification

Tables B.1– B.12 only show the overall classification accuracy rate of different algorithms. However, in specific areas of PD source identification, a knowledge of the degree of membership of a test sample to a class of data would be beneficial rather than just a class label. Such knowledge enables probabilistic interpretation of an unknown PRPD pattern that is being classified.

Of the algorithms implemented so far, Fuzzy kNN [61], Fuzzy SVM [60], and Bayesian [2, 3] have the capability to calculate posterior probability of a test sample belonging to each class of data. Besides, as shown in Tables B.1– B.12, these algorithms also have a higher classification accuracy rate. To demonstrate the posterior probability calculated by these algorithms, 7 data samples were randomly selected. The first 6 samples were from the samples that were correctly classified. The 7^{th} sample was from those that were misclassified. The probabilistic classification results for these 7 samples are shown in Tables 4.1– 4.3 where we have used the classifiers of Fuzzy kNN, Fuzzy SVM, and Bayesian. In Table 4.1, for example, FSVM/KPCA has been employed. Each of the 7 samples has a different posterior probability that shows its degree of membership to different sources of PD. In this table, sample one, which is originally from class 1, is determined to belong to classes 1 to 6 with probabilities of 84.2%, 2.4%, 0.4%, 0.1%, 12.9%, and 0.0%, respectively. Sample 7 that originally belongs to class 3, however, is misclassified to class 2 with a posterior probability of 36.5%. Its degree of membership to class 3 (the correct class) is 30.8%.

Determination of "degree of membership" for PRPD test samples would allow safer decision making considering the risk associated with different sources of PD in HV apparatus. Posterior probability level of a sample belonging to a class of data has some other advantages. One of these advantages is mostly important for the class prediction of a new unknown PRPD sample which is generated form the same type of defect but does not originally belong to the original dataset. This probability shows how similar a sample is to the class that it has been classified into, and also how much the probability of this sample belonging to other classes of data is. Based on this probability, it is even possible to reject a sample from classification by setting a threshold for an

acceptable "degree of membership." This also allows taking the risk of different PD sources into account. Such ability will, for example, require a marginal classification to be referred to an expert operator. The threshold for different classes of PD would be defined based on the risk imposed by a specific source of PD for the safe operation of HV apparatus under test.

Table 4.1: FSVM classification posterior probability rate for 7 PD test samples on data output of **KPCA**

Data Point	**True Class**	**Class 1** %	**Class 2** %	**Class 3** %	**Class 4** %	**Class 5** %	**Class 6** %	**C/M**
1	**1**	**84.2**	2.4	0.4	0.1	12.9	0.0	*C*
2	**2**	0.0	**74.0**	10.1	0.0	15.9	0.0	*C*
3	**3**	1.1	3.0	**79.6**	4.8	9.5	1.9	*C*
4	**4**	10.4	1.8	13.6	**66.3**	6.1	1.7	*C*
5	**5**	1.6	12.8	8.5	0.5	**75.9**	0.6	*C*
6	**6**	0.0	0.0	0.3	0.0	0.0	**99.7**	*C*
7	**3**	1.6	**36.5**	30.8	2.4	27.2	1.5	*M*

Sample Classified (C), and Misclassified (M). Class 1. Floating electrode in SF_6; **Class 2**. Point-plane electrodes in SF_6; **Class 3**. Free aluminum particle in SF_6; **Class 4**. Free aluminum particle in oil; **Class 5**. Point-plane electrodes in oil; **Class 6**. Point-plane electrodes in air.

Table 4.2: Bayesian classification posterior probability rate for 7 PD test samples on data output of **PCA**

Data Point	**True Class**	**Class 1** %	**Class 2** %	**Class 3** %	**Class 4** %	**Class 5** %	**Class 6** %	**C/M**
1	**1**	**100**	0.0	0.0	0.0	0.0	0.0	*C*
2	**2**	0.0	**69.5**	0.0	0.0	30.5	0.0	*C*
3	**3**	0.0	31.6	**60.6**	0.0	7.8	0.0	*C*
4	**4**	0.0	0.0	0.0	**50.9**	49.0	0.1	*C*
5	**5**	0.0	14.2	11.2	0.2	**74.5**	0.0	*C*
6	**6**	0.0	11.0	17.7	0.2	7.2	**64.0**	*C*
7	**4**	0.0	0.1	0.0	38.7	**61.1**	0.0	*M*

Table 4.3: FkNN classification posterior probability rate for 7 PD test samples on data output of **LPP**

Data Point	**True Class**	**Class 1** %	**Class 2** %	**Class 3** %	**Class 4** %	**Class 5** %	**Class 6** %	**C/M**
1	**1**	**86.3**	0.0	0.0	2.9	10.8	0.0	C
2	**2**	0.0	**77.0**	16.6	0.7	5.8	0.0	C
3	**3**	0.0	31.7	**62.6**	0.0	5.8	0.0	C
4	**4**	0.0	0.0	0.0	**86.3**	13.7	0.0	C
5	**5**	0.0	18.0	5.8	2.2	**74.1**	0.0	C
6	**6**	0.0	13.7	3.6	0.0	1.4	**81.3**	C
7	**2**	0.0	38.1	**43.9**	0.0	18.0	0.0	M

4.2 Experimental Procedure of Automated Recognition of PD Source in Oil-Immersed Insulation

To test our PRPD classification system, its application on identification of partial discharge sources in oil-immersed insulation has been investigated [7]. Three sources of partial discharge are simulated to generate artificial partial discharge data; bubble wrap to simulate air bubbles, needle to simulate corona discharge, and metal particles. Fingerprints from phase resolved partial discharge patterns are extracted. Dimension reduction techniques are employed to reduce the size of the collected data. Classifiers are developed for partial discharge source identification. A test cell is built to model three different types of PD activities with different mechanisms in oil insulating media.

The results of this subsection show promising possibility to conduct PD monitoring of power transformers which are important pieces of equipment in power systems. Because of this importance, their reliability and desirable electrical condition are required to be maintained and to this end PD detection could be an important diagnostic technique for the assessment of their liquid (oil) insulation [*e.g.* [76,77]]. In general, reliable, automated classification of PD sources in oil-immersed insulation results in online monitoring of power transformers more accurately and efficiently.

4.2.1 Test Cell Configurations in Oil-immersed Insulation

In this part, we investigate the relationship between the variation of PRPD patterns and the type of oil-immersed PD sources. Laboratory measurement tests are done on a test cell which is built to model PD activities with different mechanisms in oil. These different PD sources are briefly discussed. Artificial air bubbles are generated using bubble wraps to model small air bubbles, as a common source of PD in oil insulation. Air bubbles strongly decrease the reliability of insulation [78]. The correlation between the number of air bubbles and PRPD patterns are analyzed and are compared to the patterns which have been presented for air bubbles in the literature [76–78]. It will be shown that the number of air bubbles influences the PRPD patterns. Further to air bubbles, floating metal particles, as another common source of PD in oil-insulated HV apparatus, are studied. Moreover, in order to model point-plane discharges in oil, a needle electrode configuration is also used. The automated classification of such PD sources is performed. Classification of different sources of PD in oil requires a thorough database for training and testing. Because of costly nature of data collection in a high voltage transformer and that it requires lots of efforts, such database is generated based on the measurements conducted on small size artificial defects that are implemented in a small laboratory test cell. This test cell is built to model three different types of PD activities with different mechanisms in oil insulating media. These different PD sources are briefly discussed.

4.2.1.1 Bubble Wraps (Small Air Bubbles)

Gas bubbles as a common source of PD in oil insulation decrease the reliability of insulation because of their deleterious effect on dielectric strength of insulation. In this research, in order to model PD due to bubbles, a test cell with two electrodes with a gap of 7 mm is used. To model air bubbles in oil, bubble wraps with different number of bubbles are employed on the ground electrode. A photo of this test cell is shown in Fig. 4.11. To investigate the effects of number of bubbles on PRPD patterns, different numbers of bubbles (1, 2, 4, and 7) have been used.

Fig. 4.11: The geometry of the test cell electrodes with bubble wrap (diameter of each bubble is 7 mm.

4.2.1.2 Floating Metal Particles (Shavings)

The sources of external metal particles in oil might be transformer walls or the oil circulation system. These particles are small (with a typical length of 2 mm) and light enough to drift due to the electric filed strength or circulation of the oil. In this experiment, the electrodes have a gap of 10 mm and a holder is placed horizontally between them to keep metal shavings in between them. To avoid breakdown, both electrodes where covered by two pieces of paper. A photo of this test cell is shown in Fig. 4.12.

Fig. 4.12: The geometry of the test cell electrodes with floating metal particles (shavings).

4.2.1.3 Needle Electrode in Oil

To model point-plane discharges in oil, as another source of PD in oil, a 5 μm tungsten needle electrode configuration is used. The high voltage is connected to the needle with its tip located 10 mm away from the grounded electrode. Like previous tests, to avoid breakdown, grounded electrode is covered with paper.

4.2.2 Results and Discussions

4.2.2.1 PRPD Patterns of Test Cells in Oil-immersed Insulation

The PRPD patterns of the bubble wraps (small air bubbles) with different number of bubbles are shown in Fig. 4.13. A control experiment where all the bubbles were burst was also performed. No PD was observed for about 10 minutes running the experiment with the voltage increased to 24 kV. The patterns are almost symmetrical in each half cycle. Comparison of these PD patterns with those found in [76–78] shows a good agreement. The inception voltage for one bubble was about 11 kV but for 2 bubbles and more it increased to 16 kV. All the PRPD patterns in Fig. 4.13

are captured at a voltage of 20% above the inception voltage (at this level of voltage, PD patterns are more differentiable).

The PRPD pattern of the floating metal particles (shavings) is shown in Fig. 4.14. Before energization of the test cell, all the particles were located randomly and far apart from each other. However, after energization of the test cell, the particles started to line up in the direction of the electric field. The partial discharges started at an inception voltage of 2.8 kV. The applied voltage of this experimentis is 20% higher than the inception voltage.

The PRPD pattern of the needle electrode in oil is shown in Fig. 4.15. Inception voltage of this test cell was 20 kV. The PRPD pattern is captured at a voltage of 20% higher than the inception voltage. Consistent with the results of [78], it is observed that PD effects in this pattern are more vigorous in the positive half cycle with a large dispersion.

The measurement procedure and system calibration has been performed according to the IEC60270 standard. A coupling capacitor, transformer, measurement impedance and commercial measuring equipment (Omicron MPD600) are parts of this setup which have been used for the experiments.

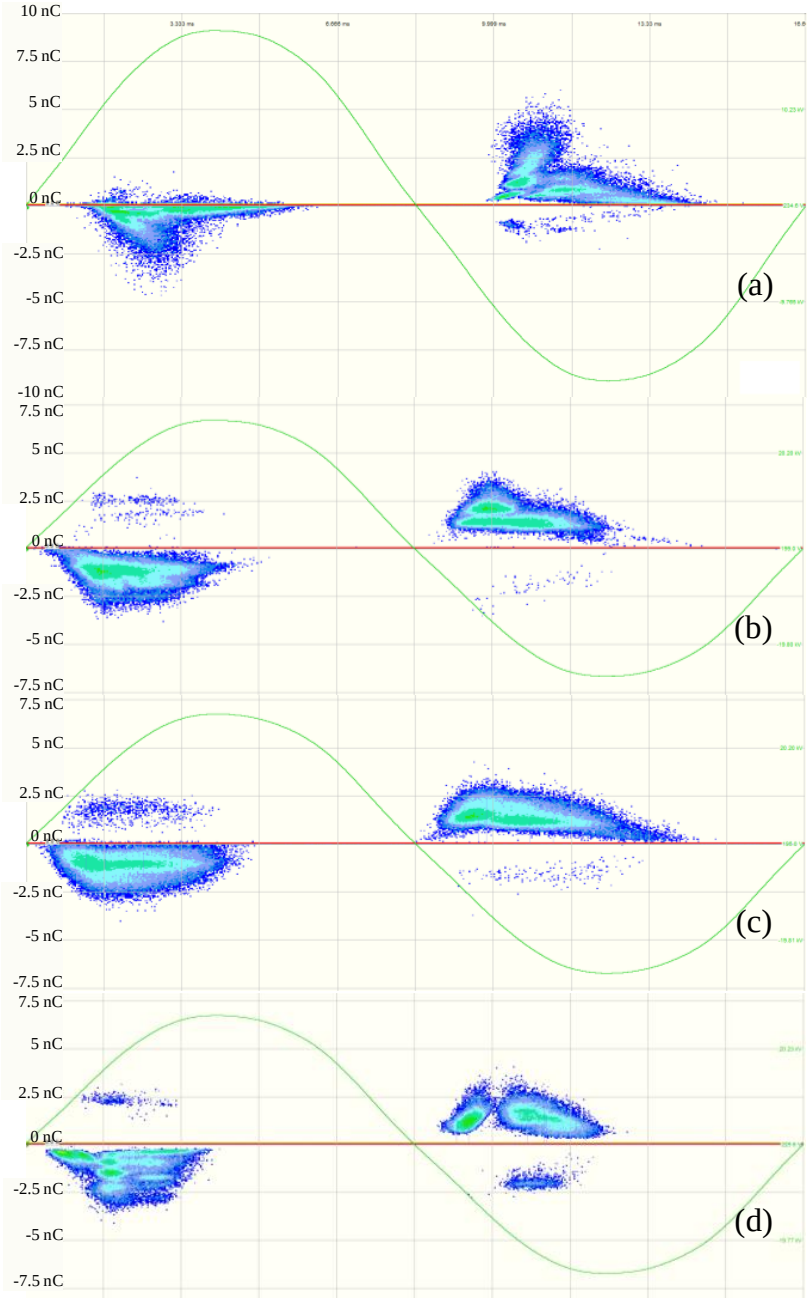

Fig. 4.13: PRPD pattern of air bubbles simulated by bubble wraps, (a) 1 bubble; (b) 2 bubbles; (c) 4 bubbles; and (d) 7 bubbles.

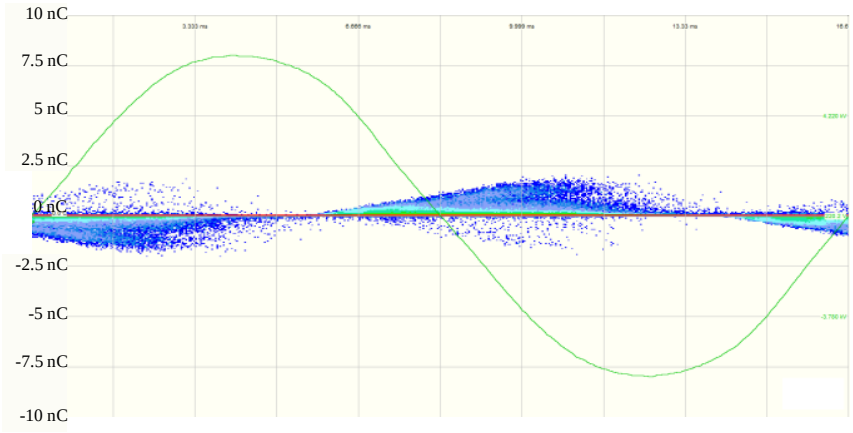

Fig. 4.14: PRPD pattern of floating metal particles (shavings).

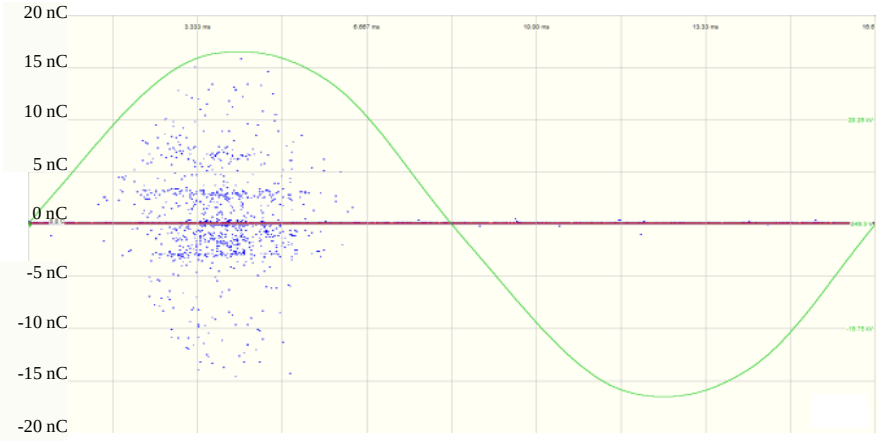

Fig. 4.15: PRPD pattern of a needle electrode.

4.2.2.2 Performance Analysis of Classifiers

Training of the classifiers is performed based on 130 data samples which were collected for each one of the PD sources to make a dataset. This dataset is used to find the classifier parameters. For testing the classifiers and for increasing the reliability of performance evaluation, each test sample is captured from the experiment which was re-performed on the same test set after several hours. In total, 30 test samples were collected for each one of the PD sources. The test confusion matrices evaluated by kNN (for $k = 4$) and SVM classifiers on the PCA and FDA results are demonstrated in Tables 4.4– 4.7. Each table presents the overall classification success rate related to the specific feature extraction/classification algorithms.

The dimension of the new dataset by PCA is equal to 7 (which corresponds to the intrinsic dimensionality of the data determined by maximum likelihood estimation) and is equal to 5 for FDA. This number is one less than the number of classes. The first three components of data from FDA and the first three principal scores of data from PCA are ploted in Fig. 4.16 and Fig. 4.17, respectively. Comparing these figures, one can see that the transformed data points with PCA which belong to different sources of PD are more scattered than those transformed using the FDA. As expected, this is due to the nature of FDA algorithm described in Chapter 3. Using both PCA and FDA, all classes of data except the classes of 2, 4, and 7 bubbles, are well separated compared to each other. This could be due to the fact that classes of 2, 4, and 7 bubbles have more similar PRPD patterns compared to that of the 1-bubble class. The misclassification of bubble PDs observed in Tables 4.4– 4.7 is consistent with the similraity of the PRPD patterns shown in Fig. 4.13. Overall, the application of SVM integrated with PCA algorithm on PD classification outperforms other three feature extraction-classification algorithms with a classification rate of 96.11%.

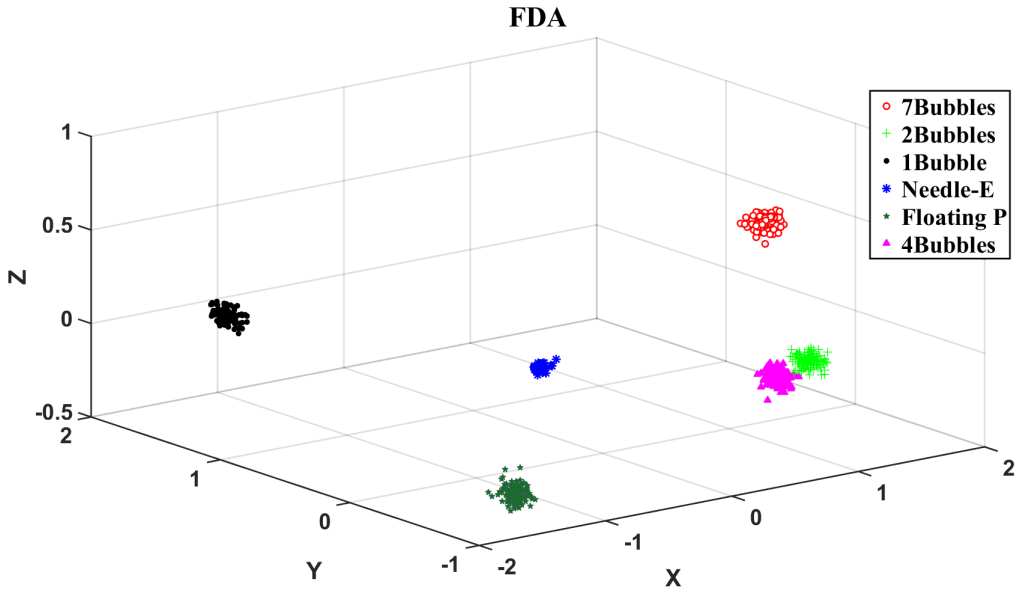

Fig. 4.16: First three components of data from FDA.

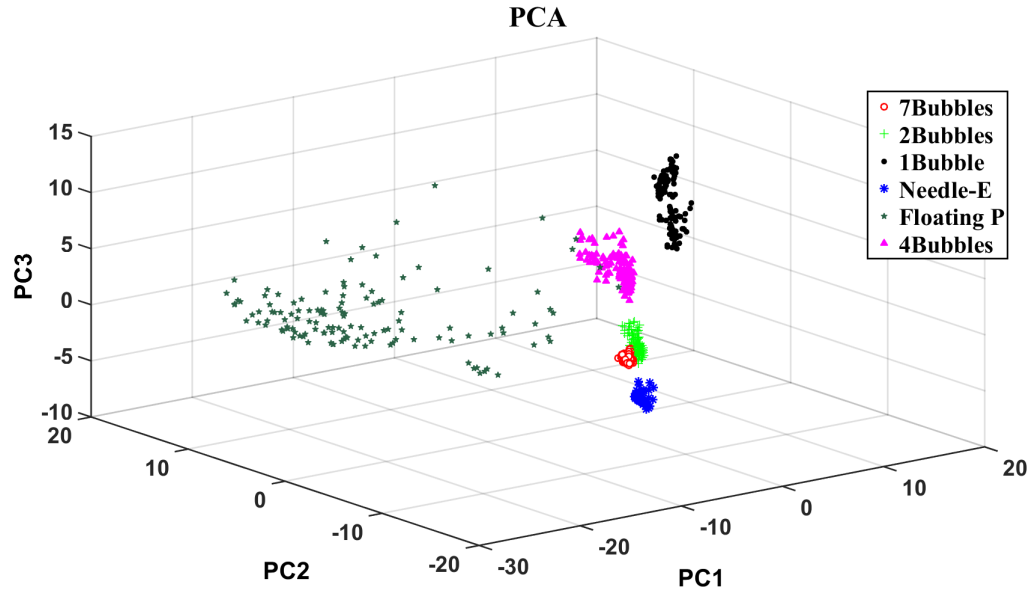

Fig. 4.17: First three principal scores of data from PCA.

Table 4.4: kNN on **PCA** results (classification rate: 92.22%)

	7B	**4B**	**2B**	**1B**	**Needle**	**Floating**
7B	16	0	14	0	0	0
4B	0	30	0	0	0	0
2B	0	0	30	0	0	0
1B	0	0	0	30	0	0
Needle	0	0	0	0	30	0
Floating	0	0	0	0	0	30

Table 4.5: SVM on **PCA** results (classification rate: 96.11%)

	7B	**4B**	**2B**	**1B**	**Needle**	**Floating**
7B	23	0	7	0	0	0
4B	0	30	0	0	0	0
2B	0	0	30	0	0	0
1B	0	0	0	30	0	0
Needle	0	0	0	0	30	0
Floating	0	0	0	0	0	30

Table 4.6: kNN on **FDA** results (classification rate: 90.56%)

	7B	**4B**	**2B**	**1B**	**Needle**	**Floating**
7B	29	0	1	0	0	0
4B	0	30	0	0	0	0
2B	0	11	19	0	0	0
1B	0	0	0	30	0	0
Needle	0	0	0	0	30	0
Floating	0	0	0	0	5	25

Table 4.7: SVM on FDA results (classification rate: 93.89%)

	7B	**4B**	**2B**	**1B**	**Needle**	**Floating**
7B	30	0	0	0	0	0
4B	0	30	0	0	0	0
2B	0	3	27	0	0	0
1B	0	0	0	30	0	0
Needle	1	0	0	0	29	0
Floating	0	0	0	2	5	23

4.3 Experimental Procedure of Automated Recognition for Thermally-degraded Cellulose-oil Insulation

The third test setup is designed to show the performance of the proposed classification system on identification of PD sources in power transformer cellulose insulation samples while they are under both electrical and thermal stresses [8]. In this part, to collect necessary information for making a thorough dataset, laboratory measurement tests are performed on a test cell which is built to model surface discharges (tracking) at various temperatures on the interface of pressboard-oil insulation which are one of the major causes for the failure of transformer insulation [79].

The importance of this part of the research is due to the fact that, the reliability of transformers mainly depends on proper insulation design [80] and existence of cellulose materials[6] as part of the insulation system of transformers is to prevent the emission of electronic charge that will occur from conducting surfaces. However, these materials age due to the heat generated by the core and windings during normal operation. Over time, aging of the cellulose insulation causes the loss of tensile strength. If this deterioration continues, the cellulose materials become more brittle which in case of system disturbances, such as a through fault or load rejection, results in most transformer failures. Aging of the cellulosic insulation will also be accelerated by water. Moisture inside the solid insulation decreases their reliability. Moisture is harmful for cellulose insulation. It can accelerate

[6]Cellulose insulation used in pressboard is a light, fibrous, and porous material.

the cellulose rate of aging and reduce both dielectric and mechanical strength of cellulose insulation. Reliable, automated analysis of PD sources in different level of temperature proves the effects of temperature on phase resolved partial discharge patterns. We have also tried to use the proposed classification system to investigate the correlation between the variation of specific features in PRPD patterns and the level of working temperature. To perform that, laboratory measurement tests are performed on cellulose-oil insulation under simultaneous thermal and electrical stresses. The samples are energized in an oven at different levels of temperature to show the dependency of partial discharges on working temperature.

Based on the results which are shown in follows, it is concluded that the presence of the moisture is an important parameter in development of surface discharges (tracking process). Temperature also is important because it affects strongly the moisture transition from pressboard to oil or vice versa. Based on this, we speculated that PRPD pattern differences that exists are due to the changes in moisture level and physical structure of the samples when they are simultaneously heated and electrically energized. This might be due to the dependency of distribution of moisture in oil and pressboard to temperature. The recognition results show successful performance in this area of studies and to some extent indicate the promising possibility of online and offline automatic classification of PD sources in the cellulose-oil insulation of power transformers. This capability enables online monitoring of high voltage cellulose insulation more accurately and efficiently which helps to prevent most transformer failures.

4.3.1 Test Cell Configurations in Thermally-degraded Cellulose-oil Insulation

To produce surface discharges on the interface of pressboard-oil, a dimpled cellulose pressboard with 1 mm thickness was used. Each of pressboard samples were approximately 100 mm by 80 mm. A stainless steel needle with a 20 μm tip radius was used for the high voltage (HV) electrode which is placed at an acute angle (approximately 30°) to the pressboard at a distance of 30 mm from an aluminum bar (10 mm×10 mm×80 mm) for the ground electrode which is placed on the surface of

the pressboard [81]. Electrodes and pressboard are immersed in mineral oil and the whole setup is placed in an oven (see Fig. 4.18). Placing the needle in the acute angle is useful to direct the charge from needle to ground bar along the surface of the pressboard [81]. A test transformer was used to energize this arrangement and an oven is used to generate three levels of temperature which were $25°C$, $90°C$, and $110°C$. Partial discharges measured using a 1 nF coupling capacitor along with an Omicron MPD600 partial discharge measuring system. The applied voltage was increased to a level between $15 - 25$ kV on each of the pressboard samples at different temperatures. This was sufficient to produce regular discharges with pulse magnitudes greater than 1 nC. The samples subjected to electrical stress were exposed to maximum of 2-hours of surface PD.

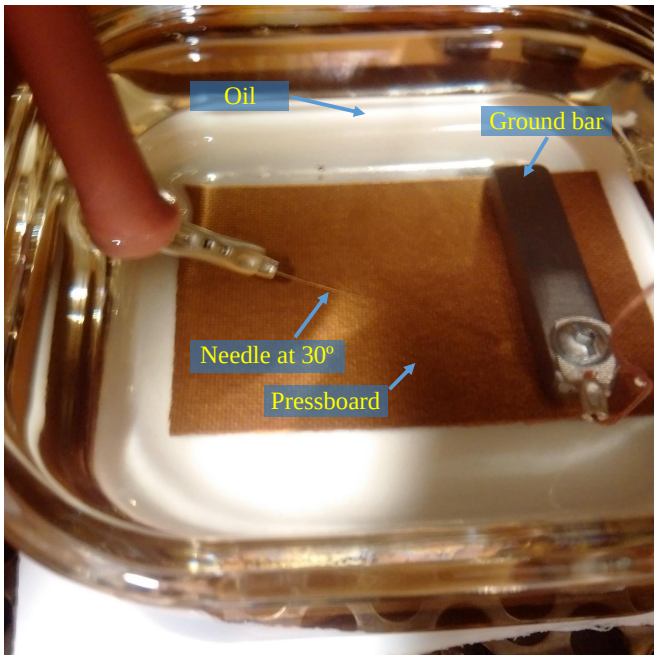

Fig. 4.18: Needle-bar electrode test arrangement used to produce surface PD.

4.3.2 Results and Discussions

4.3.2.1 PRPD Patterns of the Test Cell Over a Temperature Trend

Phase-resolved PD diagram for the samples were captured while under simultaneous stresses. These samples are energized at normal room temperature $25°C$, $90°C$ and $110°C$. The samples in different temperature levels of $25°C$, $90°C$, and $110°C$ had similar inception voltages around $19-20\,\text{kV}$. The AC test voltage was increased to a level of 25 kV in order to produce surface discharges exceeding 1 nC in magnitude. At room temperature ($25°C$), the PD pulse magnitudes were typically between $1-2\,\text{nC}$. When temperature was increased to $90°C$, the PD pulse magnitudes increased slightly to about $1-2.5\,\text{nC}$ and at the $110°C$ PD pulse magnitudes were typically increased to $1-8.5\,\text{nC}$. The occurrence of discharges was increased from $25°C$ to $110°C$. All PD patterns were asymmetrical with larger discharge pulses on the positive half-cycle than the negative half-cycle that is common for surface discharges [79]. In Fig. 4.19 and Fig. 4.20, the PD patterns of surface discharges for two different temperatures of $25°C$ and $110°C$ are plotted.

After capturing the PRPD patterns, the results of automated recognition system based on application of two feature extraction algorithms which are PCA and FDA are demonstrated. These two algorithm are applied on the resulting patterns to reduce the dimensionality of the data and to prepare them to be efficiently distinguished.

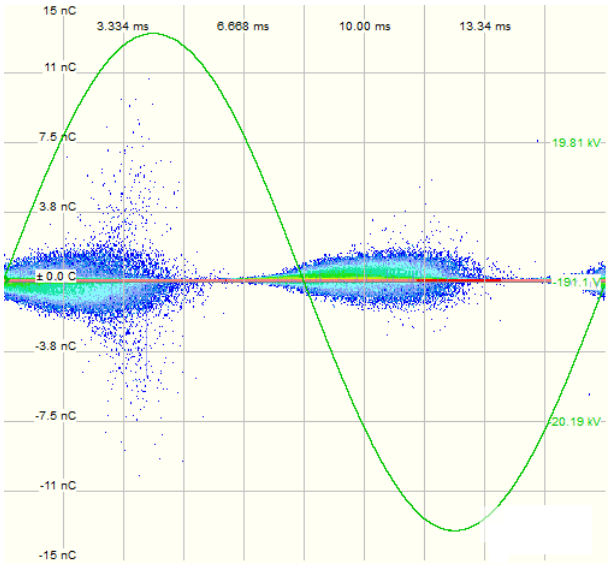

Fig. 4.19: PRPD pattern of surface discharges on the interface of pressboard-oil insulation in $25°C$.

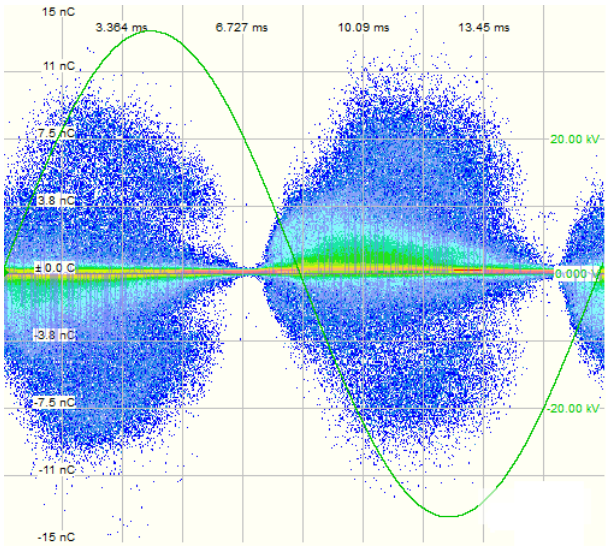

Fig. 4.20: PRPD pattern of surface discharges on the interface of pressboard-oil insulation in $110°C$.

4.3.2.2 Performance Analysis of Measurements and Classifications

At room temperature, active surface discharges on the pressboard resulted in a small carbon tracks near the needle tip and gas channels and white marks in the pressboard which branch out in a tree-like fashion. Formation of white marks at the pressboard surface had occurred during the testing. These white marks were not permanent (see Fig. 4.21) and after the system was de-energized the white marks were gone with oil leaving no visual indication that discharges had occurred in these regions. However, at $90°C$, these white marks in the pressboard were smaller but carbon tracks near the needle tip got bigger. The system was energized for two hours at $25°C$ and $90°C$ and no fault (surface flashover) happened. At $110°C$, there was no sign of white marks in the pressboard while carbon tracks were bigger and directed toward the ground bar. After 11 minutes a fault (surface flashover) happened. When fault happened, ground bar moved toward the needle and under ground bar, some big carbon spots could be easily seen (Fig. 4.22). Surface PD data shows clear correlation to the level of temperature. Partial discharge inception occurred at the same voltage level (19.5kV), but the magnitude of the PD was much larger at $110°C$ than for other samples and the flashover happened at this temperature. A large carbon spot formed near the needle tip, and carbon spots also formed near the ground electrode. These differences could be attributed to the increase of moisture in the oil with the increase of temperature because as distribution of moisture in oil and pressboard is a function of temperature [82].

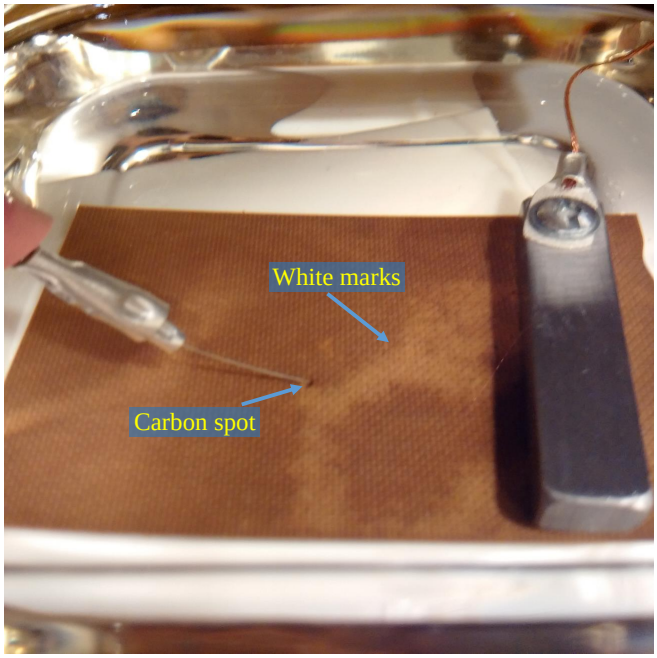

Fig. 4.21: Carbon tracks and white marks on the pressboard at $25°C$.

Fig. 4.22: Carbon tracks on the pressboard at $110°C$ after fault occurred.

The justification for these behavioral differences in higher temperatures could be explained by the point that when the temperature increases, moisture is generated in the pressboard and released. Part of it even remains in the pressboard and moves to and stays mostly in pressboard surface. As a result, the water contents in insulating oil and surface of the pressboard increases. Also, the chemical reactions of pressboard degradation and oil oxidation provide water as a by-product [83]. Water generally has low solubility in transformer oil. With an increase in temperature, the water solubility in oil can be significantly increased [84] . On the other hand, free water can be formed in oil. So moisture also enters from the environment and heated oil can absorb that water and contain some amount of water. The results of the investigations in the past have also confirmed the reduction of breakdown voltage in transformer oil during moisture transition from pressboard to oil or vice versa [82]. In the literature, it has also been shown that the electrical strength depends mainly on the water content in pressboard and the gradient of temperature [82]. As the moisture increases in oil and pressboard surface by temperature, experimental results makes it clear that the total charge transfer, due to the surface discharges, is increased. Another theory which is worth to be mentioned is the relation of the total charge transfer, due to the partial discharges, by temperature. This happens because of the effects of temperature on the conductivity of the dielectrics. The experiment shows that partial discharges depend on the temperature. To understand these differences between surface discharge pulses in different temperatures and at the same time showing the promising probability of automated recognition of such PD sources, captured PRPD patterns for different temperatures have been analyzed.

The partial discharge data recorded for 3 seconds (180 cycles of the applied voltage) generate one data sample with a dimension of 300 (explained in Chapter 3). In each experiment, 120 data samples have been saved. The dimension of the new dataset by PCA is equal to intrinsic dimensionality of the data determined by maximum likelihood estimation, however, for visualization the first three dimensions have been plotted which captures most percentage of the variance. The visualized dimensions by FDA is equal to 2, that is one less than the number of classes ($25°C$,

$90°C$, and $110°C$). The first two components of data from FDA and the first three principal scores of data from PCA are plotted in Figs. 4.23 and 4.24, respectively. Comparing these figures, one can see that the transformed data points with PCA which belong to the PD sources at different temperature levels are more scattered than those transformed using the FDA. As expected, this is due to the nature of FDA algorithm described in Chapter 3. Using both PCA and FDA, all classes of data are well separable.

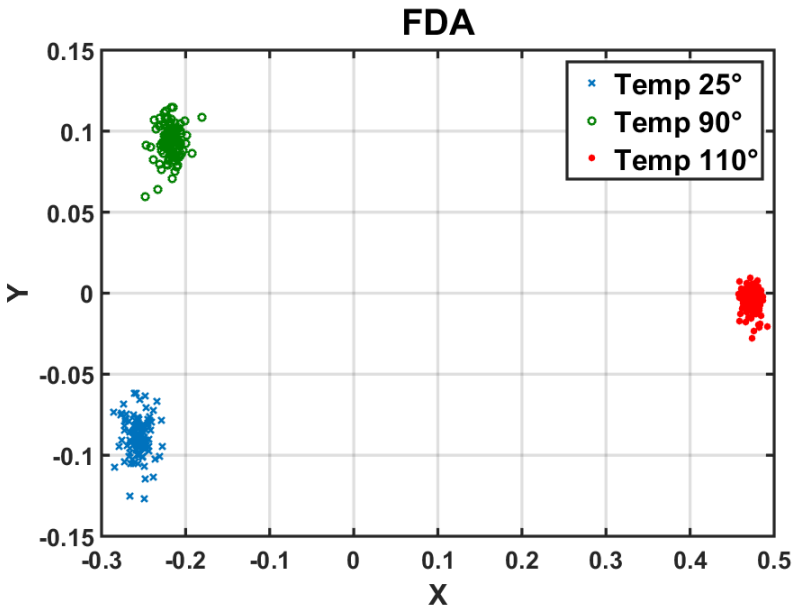

Fig. 4.23: Two components of data from FDA.

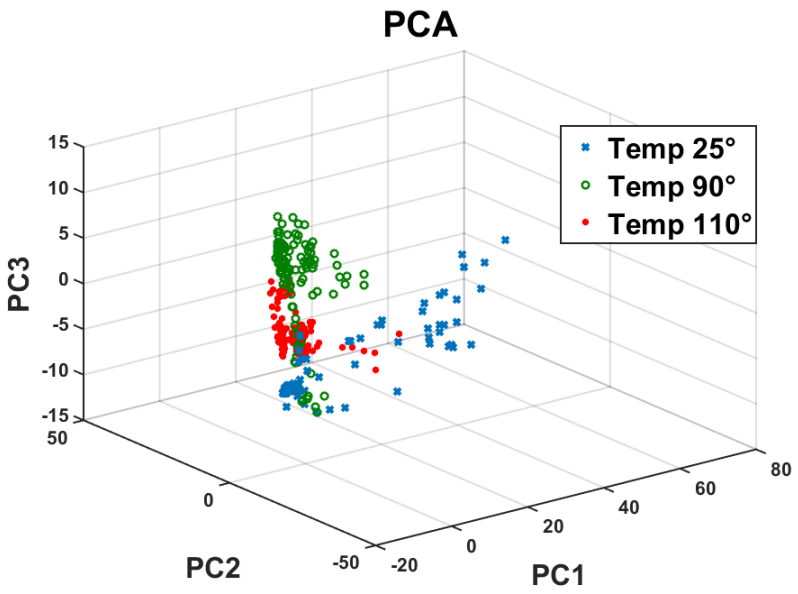

Fig. 4.24: First three principal scores of data from PCA.

4.4 Summary

In this chapter, the implementation and performance evaluation of the designed PD source identification system using PRPD patterns of different single sources of PDs were presented. Different artificial laboratory test subsets are designed and built to model PD sources in a variety of insulation media and also different working temperatures.

The first test setup includes test cells which are designed to model common sources of PD in air, oil, and SF_6. In this part of my PhD thesis, the application of automated classification system on different sources of PD in different HV insulation media was investigated. The laboratory measurement tests were performed on test cells that are built to model PD activities with different mechanisms in air, oil, or SF_6. These test cells include 3 sources of partial discharge in SF_6 (floating electrode, moving particles, and fixed protrusion), 2 sources of PD in transformer oil (free particle and needle electrode), and corona in air. Eventually, high classification accuracy rate

results of automated classification system on PD sources of insulation based on different feature extraction and classification algorithms were demonstrated. The results showed that FSVM and KSVM integrated with MDS outperform other feature extraction-classification algorithms with a classification rates of 99.4% and 99.1%, respectively. However, application of classifier algorithms on MDS, KPCA, Isomap, SNE, and LPP show high accuracy classification rate. From these results, it could be concluded that not only the nonlinear feature extraction algorithms work properly when applied on PD datasets, but also some of them outperform the classification results by linear algorithms and statistical operators. Classification using SPE does not show any desirable accuracy compared to other feature extraction algorithms. Classifier algorithms integrated with PCA and FDA from the traditional linear group show acceptable performance and they even work with higher classification accuracy compared to statistical operators. Probabilistic interpretation of an unknown PRPD pattern that is to be classified was presented using some of the applied classifier algorithms. These classifier algorithms, including Fuzzy classifiers (FSVM, FkNN) and Bayesian, are able to show a high accuracy rate of classification further to providing a knowledge of the "degree of membership" of a test sample to a class of data. This could be more beneficial rather than a class label assignment. Such knowledge enables probabilistic interpretation of an unknown PRPD pattern that is being classified. Overall, these classification results and availability of posterior probability show prosperous performance in this area of studies and to some extent indicate the promising possibility of online and offline automatic classification of PD sources in HV apparatus.

Second test setup is designed to test the classification system on the identification of different sources of PD in oil-immersed insulation. These sources of partial discharge are simulated to model common PD sources in HV transformer which includes, bubble wrap to model small air bubbles, needle electrode in oil to simulate point-plane discharge, and floating metal particles (shavings) in oil. The laboratory experiments were performed on the test sets which are built to model PD activities with different mechanisms in oil. The results of automated classification system on oil-immersed insulation PD sources based on two integrated feature extraction and classification algorithms were

demonstrated. These results show prosperous performance of the proposed automatic classification system on the identification of different sources of PD in oil-immersed insulation. This achievement could help in the area of monitoring and automatic classification of PD sources in power transformers.

The third test setup is designed to show the performance of proposed classification system on the power transformer cellulose insulation samples under both electrical and thermal stresses. The possibility of automated recognition of partial discharge sources for thermally-degraded cellulose-oil insulation was investigated. The laboratory measurement tests are performed on the test sets which are built to model surface PD activities in different temperatures on the interface of cellulose-oil insulation. The effects of temperature on the development of surface discharges in the interface of the cellulose-oil insulation also was reported. It is concluded that the presence of the moisture is an important parameter in development of surface discharges (tracking process). Temperature also is important because it affects strongly the moisture transition from pressboard to oil or vice versa. The results of this part proves the existence of a correlation between the variation of specific features in surface discharge patterns and the level of working temperature. These results imply how successful the proposed automatic classification system is on the identification of different surface discharges in high voltage cellulose insulation occurring in different temperatures. This achievement similar to the results achieved from the second test setup could be helpful in the area of monitoring and automatic classification of PD sources in power transformers which aiming to prevention of most transformer failures[7].

The results show that the proposed classification system is well able to successfully identify the sources of partial discharge in different insulation media under different working temperatures. Availability of this classification system enables continuous 24/7 monitoring of equipment and helps to identify PD sources in early stages which leads to safe operation of HV apparatus. Also providing

[7]Application of the proposed automatic classification system on artificial laboratory setups shows prosperous performance in this area of studies. However, the proposed automated system needs to be validated on prototype system components, like model transformers, to make it ready for application in grid monitoring.

the probabilistic interpretation based on the risk associated with different PD sources, a marginal classified PRPD sample by the proposed classification system will be referred to an expert operator to do a visual inspection and make a proper decision; Otherwise, decision will be made by the classification system.

In online condition assessment monitoring of high voltage (HV) insulators, beside single PD source identification, it is sometimes required to identify multiple, simultaneously activated partial discharge (PD) sources that happen in the insulation of the HV apparatus. To further enhance the proposed classification system, in the next chapter, identification of PRPD patterns that are a mix of multiple, simultaneous PD sources will be performed. To do so, we develop a novel algorithm to identify multiple, simultaneously activated PD sources using PRPD patterns that are widely used in power industry and are easier to analyze compared to time-consuming PD pulse waveforms analysis which has been used by other researchers.

Chapter 5

Multiple Concurrent PD Sources Identification Using PRPD Pattern

In previous chapters, a powerful comprehensive classification system for single-source PDs using their corresponding PRPD patterns has been proposed. However, there are many practical situations where the interest lies in the identification of multiple, simultaneously activated PD sources in insulation [39]. Recently, identification of these types of defects is receiving more attention [39–43]. To enhance our PD classification system, these multi-source PDs need to be successfully classified. However, PRPD patterns associated with multiple simultaneously activated PD sources are often partially overlapped [39, 44] and this makes them very hard to be appropriately classified using available methods presented in Chapter 4. A few studies have been conducted in this regard, which are mainly based on analyzing the PD pulse waveforms attempting to separate individual concurrent pulse sources [9, 39–43]. Classification of these types of multi-source PDs is subsequently performed on each selected single-source PD using its related PRPD sub-patterns. This is usually done under the assumption that there exists a relationship between the nature of PD sources and their generated pulse waveforms, which helps distinguishing different pulse waveforms originated from different sources.

In this research, we develop a novel algorithm to identify multiple, simultaneously activated PD sources using PRPD patterns that are widely used in power industry and are easier to analyze compared to PD pulse waveform analysis. The multi-source PRPD pattern classification is developed using training and test databases that are generated from fingerprints of single-source PD patterns and probabilistic interpretation is performed following a novel two-step Logistic Regression (LR) algorithm [6]. This two-step LR algorithm is trained on the database derived from single-source patterns. The algorithm is then tested on samples that are generated with multi-source PRPD patterns. Classification of new samples is performed after passing them through a one-class kernel support vector machine (KSVM) classification algorithm [85] to distinguish those with multi-source PDs from those with single sources. Notably, the feature spaces associated with training and testing datasets are made by generating suitable features (fingerprints) from PRPD patterns. This is then followed by the application of PCA which is a linear feature extraction (dimensionality reduction) technique [48] that was explained in Chapter 3.

To evaluate the performance of our proposed algorithm for classification of multi-source patterns, PD measurements on a number of multi-source models are conducted. These artificial models are built to simulate common defects of GIS in small-scale laboratory test cells with realistic SF_6 gas condition. Comprehensive performance evaluation of this algorithm is conducted and the development of analytical equations is presented. This would help to perform classification in future using these analytical equations. Further, an important problem in PD source identification is to assign "degrees of membership" to multi-source PRPD patterns associated with each class label. This enables probabilistic interpretation of a new multi-source sample that is being classified. The availability of this degree of membership for future PRPD samples would allow safer decision-making by considering the risk associated with different sources of PDs in HV apparatus. The results of this work show capability to design a solid basis for an automated multi-source classification system and facilitate PD source identification in early stages. This could be of a great help in safety augmentation of HV apparatus, such as transformers, electric machines, cables, and GIS.

5.1 Experimental Procedure for Multiple Concurrent PD Sources Identification

To construct the necessary models in our proposed multi-source classification algorithm, different sources of PDs are required to generate data samples that are essential for training and testing the classifier. To this end, we generate databases using PD measurements that are conducted on artificial defects and implemented in small laboratory test cells. The measurement procedure and system calibration are performed according to the IEC60270 standard [86].

Figure. 5.1 shows the experimental setup that consists of an HV transformer for simultaneously energizing the multiple test cells, a capacitive voltage divider, a coupling capacitor, the quadruple impedance, the commercial PD measuring equipment (Omicron MPD600), and a PC. By changing the distance between electrodes in each test cell, PD inception voltage of that cell is adjusted. Applied voltage level ,however, was set higher than the highest inception voltage of test cells in a specific combination of cells when they were simultaneously activated. This voltage varied in the range of 10% to 50% above the biggest inception voltage when PD data acquisition was carried out.

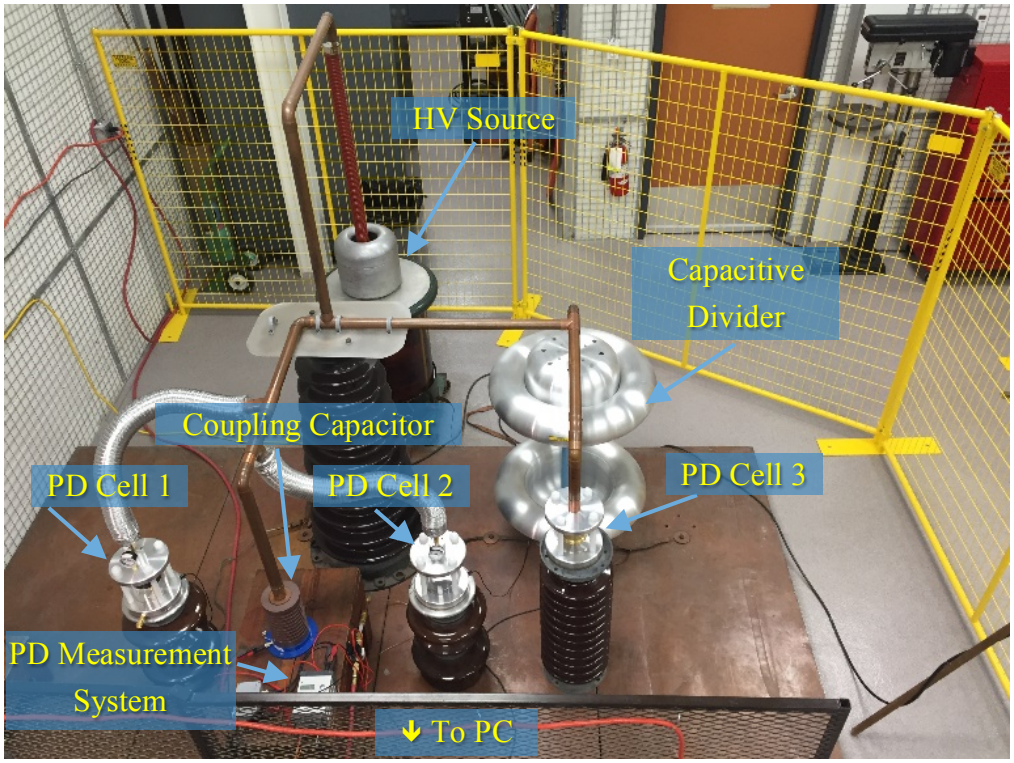

Fig. 5.1: Experimental setup of multi-source PDs (moving particle, fixed protrusion in SF_6 and fixed protrusion in air).

5.2 GIS Laboratory PD Test Cell Models and Their PRPD Patterns

To evaluate the performance of our proposed algorithm, following Hampton and Meats [70], we simulate common defects of GIS and model their corresponding PD source types in small-scale laboratory test cells as shown in Fig. 5.2. These tests cells are designed to tolerate the realistic SF_6 gas condition of GIS.

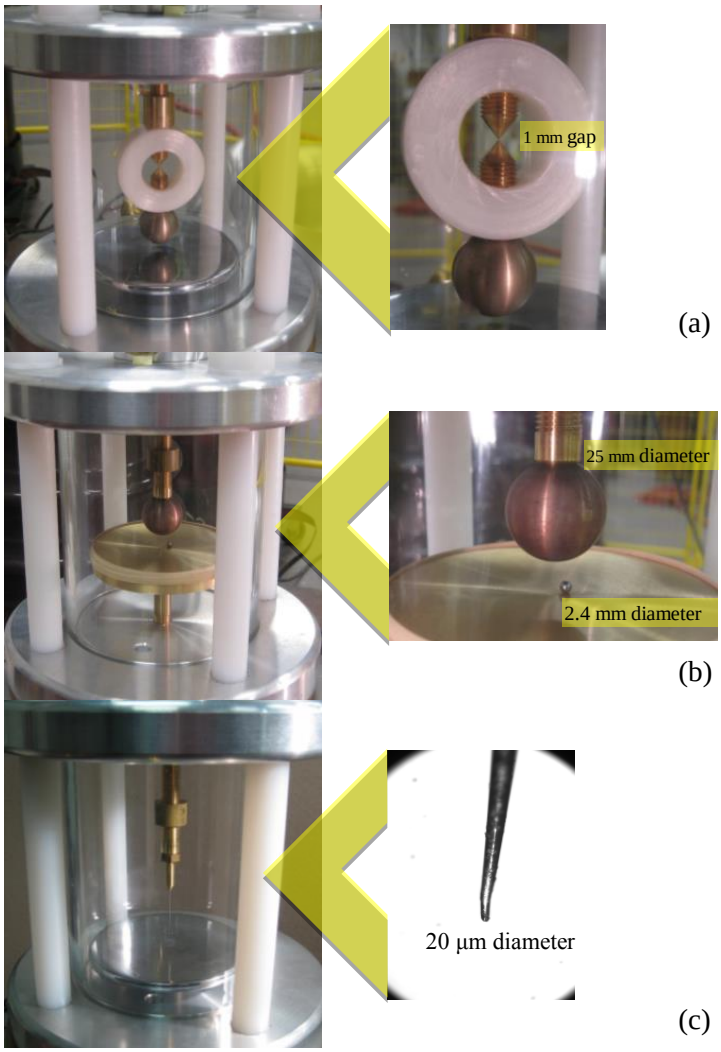

Fig. 5.2: SF_6 test cells; (a) floating electrode; (b) free particle; (c) point-plane electrodes. Each cell consists of a Perspex tube clamped by nylon screws between top and bottom aluminum caps that can withstand a pressure of up to 500 kPa.

Sparking from a floating electrode, moving particles, and fixed protrusion are some of the major sources of PDs in a GIS [71,87] whose laboratory models are shown in Figs. 5.2.a, 5.2.b, and 5.2.c, respectively. Similar to the test setups in the previous chapter, the same setup as fixed protrusion in SF_6, however filled with air at 100 kPa is also employed to generate PD due to corona in air.

These four artificial PD models corresponding to common defects in GIS, namely, moving particle, fixed protrusion, floating electrode as well as fixed protrusion in air are labeled as M, P, F and C, respectively. For F, an inception voltage of approximately 15 kV was measured at 400 kPa. We observed that both the inception voltage and PD magnitudes are strongly dependent on the gap size between the energized electrode and the floating electrode, but not much sensitive to the distance between the sphere and the ground electrode (5 mm in this experiment).

Source M includes a small bearing with a diameter of 2.40 mm located on a concave dish ground electrode. The HV electrode is a 25.4 mm diameter sphere fixed at 10 mm from the ground electrode. SF_6 pressure in this model is also set to 400 kPa and the inception voltage is equal to 15.5 kV. Source P is energized at 400 kPa and to generate PD, a tungsten needle with a tip diameter of 20 μm located at a distance of 15 mm from the ground plate was used. Inception voltage for this model is measured approximately 14.5 kV. Model C is related to corona in air. This setup is similar to the one used for the generation of corona in SF_6 but with different distance between needle and ground plate (50 mm). This experiment is done in air at 100 kPa. The inception voltage of this test cell was 14 kV. The inception voltages of all test cells remain approximately the same as those that are individually energized.

Three two-test-cell combinations and a three-test-cell combination were used to generate multi-source PRPD patterns and conduct the necessary measurements. Since the applied voltage was set higher than the inception voltages of all cells in each combination, PRPD patterns are mixed of all defects. Three-dimensional $\phi - q - n$ PD pattern of individual and combined cells are shown in Fig. 5.3.

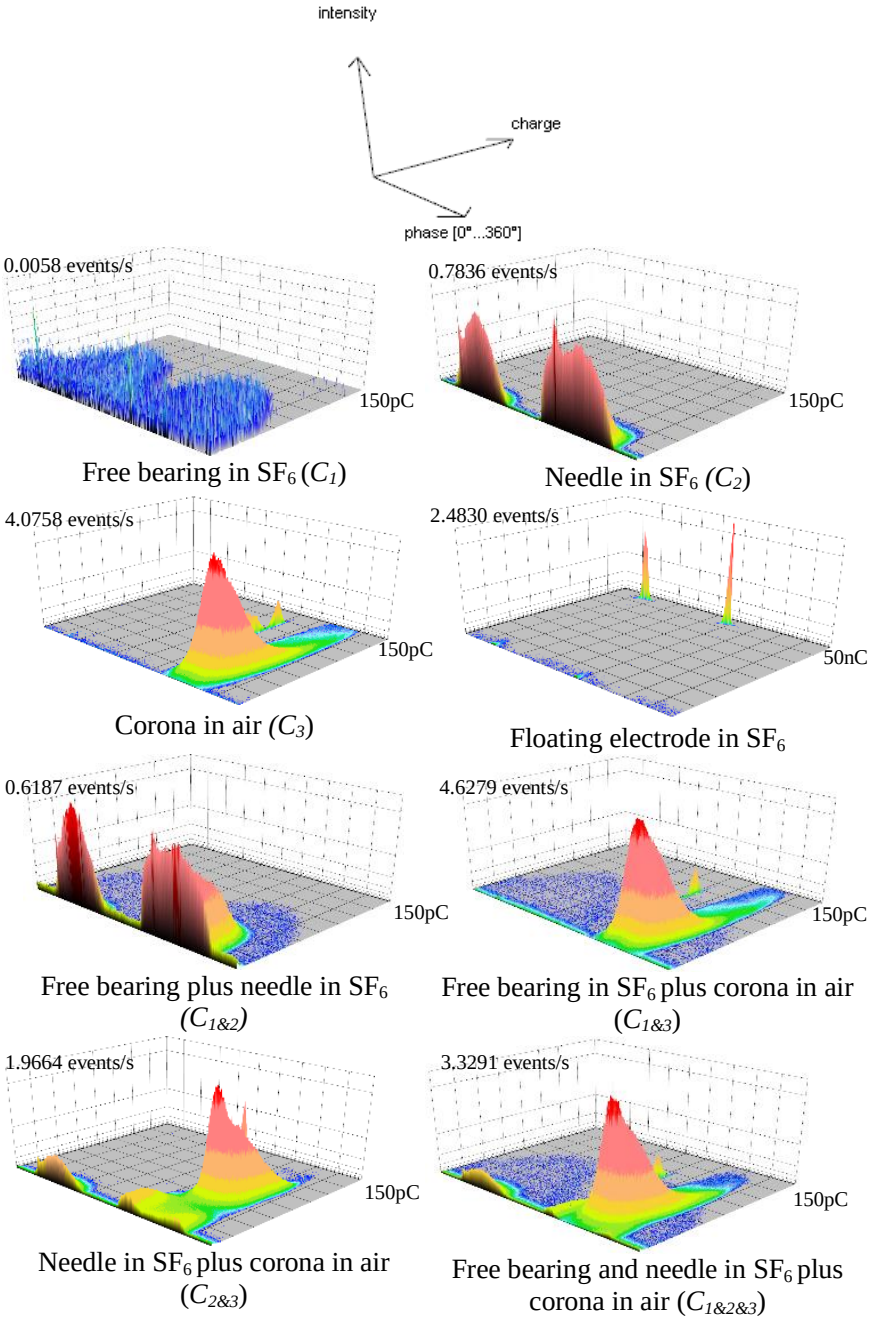

Fig. 5.3: Typical 3D "$\phi - q - n$" PD patterns of different individual cells and cells combination models.

5.3 The Proposed Algorithm for Simultaneous Multiple PD Sources Classification

In this section, we present our proposed algorithm for multi-source PD classification that may simultaneously occur in HV equipment. This algorithm is mainly composed of two parts; training and testing. Each of these two parts includes different components, whose functions were separately elaborated in Chapter 3, except for One-class SVM that will be explained in subsection 5.3.1. To construct the training part of the algorithm, PRPD patterns of single and multi-source PDs have been recorded. To generate the training samples, a novel feature generation step (which was explained in Chapter 3) is applied on each recorded PRPD pattern to represent it by its well-discriminative fingerprints. These fingerprints are generated based on the application of q-quantiles of the magnitudes of PDs in a typical PRPD pattern. This step forms data samples that are generated from three models M, P, and C. These samples are categorized in three single source classes C_1, C_2, and C_3, respectively, and later used for the training part. In addition, different combinations of single-source PD models are used to generate samples for multi-source classes $C_{1\&2}$, $C_{1\&3}$, $C_{2\&3}$, and $C_{1\&2\&3}$, which $C_{1\&2}$, *e.g.* consists of single sources of C_1 and C_2 when they are simultaneously activated.

Following the feature generation step, a feature extraction step using the PCA algorithm [48] is applied on single-source dataset to transform it from its high dimensional feature space to a new low dimensional space. To this end, single-source data samples are passed through a linear mapping which is performed by projecting data onto a new low dimensional space using a mapping matrix U that is constructed by eigenvectors associated with leading eigenvalues of the covariance matrix of the single source data. This is performed (linearly) by a simple multiplication of samples and U.

Now the purpose is to optimize regression coefficients of the first step LR model that is applied on single-source dataset. This optimization results in simple analytical formula for posterior probabilities by which the data samples belong to individual single source classes. These analytical

formulas are formed on the basis of all variables in the mapped space (eigenvectors that formed U). The second step LR model is developed to find the regression function that exists over the posterior probabilities associated with single-source PDs in the first LR model. The regression coefficients of this function are optimized upon the application of a limited number of available multi-source samples in the training set. In fact, the second LR model is employed to estimate the posterior probabilities by which a sample belongs to a multi-source PD class as a function of the posterior probabilities of the same sample belonging to single source classes. This is performed by adding a small number of multi-source samples in the algorithm, as the input to the first LR model, to find the posterior probabilities that they belong to single-source classes. These probabilities, after passing them through a PCA algorithm, are used as the input to the second LR model to optimize its regression coefficients. Only a limited number of multi-source training samples are required for the optimization process of the second LR model. This is because the second LR model is applied to posterior probabilities derived by the first LR model. The number of these variables is small and it is indeed equal to the number of single source classes available in training set. Hence, one only needs to estimate a small number of coefficients that require a limited number of multi-source samples. This is particularly important, as in practice access to training data associated with multi-source PDs could be very limited.

The testing part of the algorithm aims to classify an unknown multi-source pattern (with samples from $C_{1\&2}$, $C_{1\&3}$, $C_{2\&3}$, and $C_{1\&2\&3}$) based on the LR models established by the single source patterns template (samples of C_1, C_2, and C_3). To provide samples for testing the algorithm, same features as those used in the training set are generated from multi-source PRPD patterns. These multi-source samples are projected on the low dimensional space using the mapping matrix U (derived in the training part). To make sure that test samples belong to one of the multi-source classes, single-source classes C_1, C_2, and C_3 are considered as a unit class and a one-class KSVM algorithm [85] is developed to distinguish the multi-source test samples from single-source samples that are available in the training set. This stage aims to recognize whether the

unknown test sample belongs to one or none of the C_1, C_2, or C_3 classes. If the unknown sample belongs to one of the single-source classes then it will be passed to the traditional multiclass KSVM or LR algorithms to be classified to its corresponding class. Otherwise, it will be passed to our two-step LR algorithm. Unknown multi-source samples are entered to the probabilistic analytical equations, and then posterior probabilities of samples belonging to single-source classes are calculated. These probabilities, after implementing a PCA on them, are passed to the second set of probabilistic analytical equations to estimate posterior probabilities of the unknown samples belonging to different multi-source classes. The main advantage of this approach is the availability of analytical equations that are derived in the training part of the algorithm. Different blocks of the proposed algorithm are explained in the following subsections. The flowchart of the algorithm that was presented in this section is shown in Fig. 5.4. Performance evaluation of the algorithm shows that it works very well in identifying multi-source PD patterns based on the probabilistic relationship with their constitutive individual single source PDs.

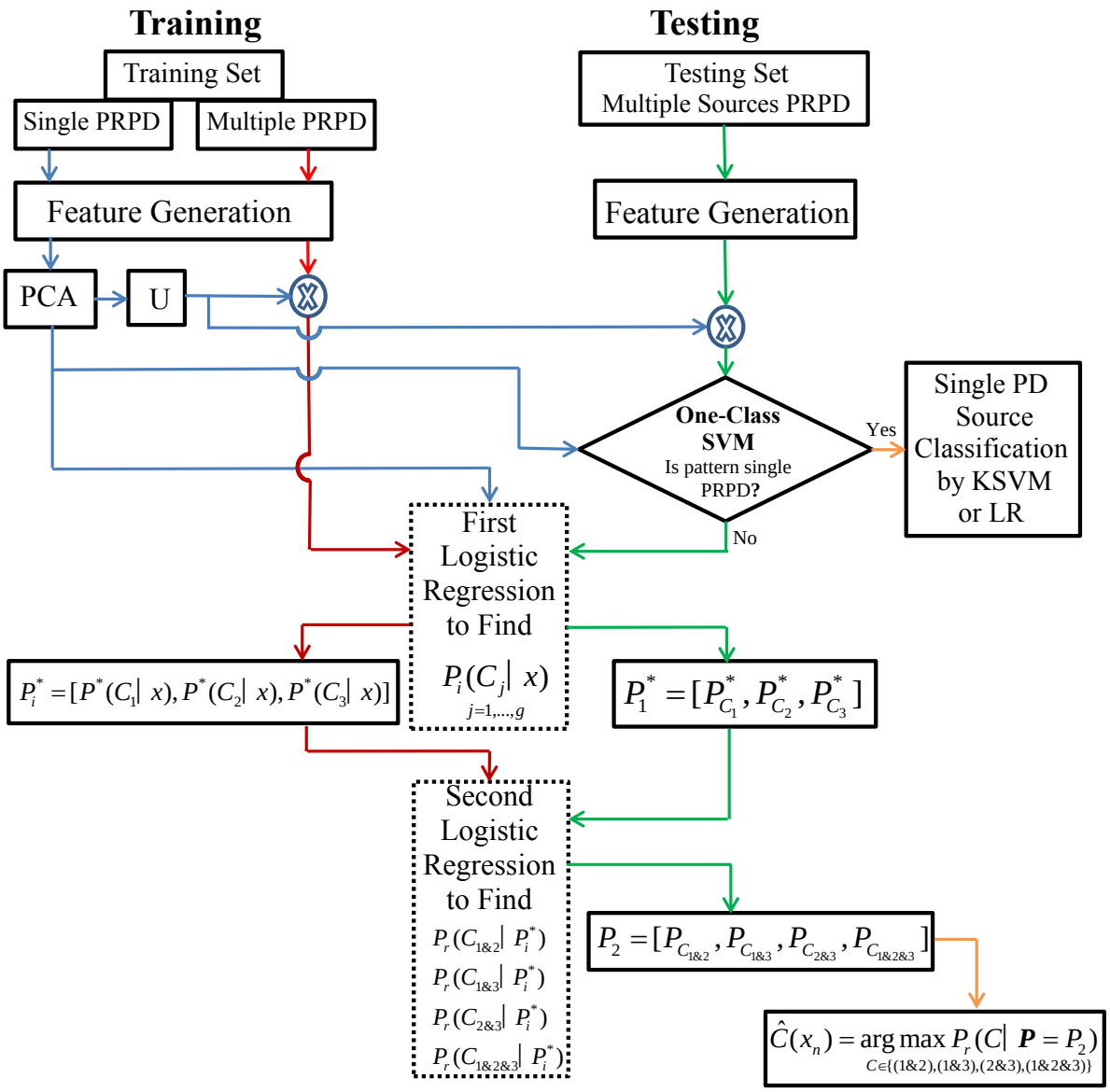

Fig. 5.4: Flowchart of the proposed algorithm- P^*s are the output of first LR model which are passed through a PCA algorithm to make them uncorrelated (independent) and appropriate as the input variables of the second LR model. g is equal to the number of single PD source classes.

5.3.1 One-class SVM

The One-Class SVM algorithm is an extension of the support vector machine algorithm [85]. This algorithm starts with mapping the data points with a feature map $\phi(x)$ into a feature space H corresponding to the kernel function, K, and then separating them from the origin with maximum margin. The quadratic programming which needs to be solved is slightly different than the main SVM cost function and it is given by [85]

$$\underset{\omega,\xi_i,\rho}{min}(\frac{1}{2}\left\|\omega\right\|^2+\frac{1}{\upsilon N}\sum_{i=1}^{N}\xi_i-\rho) \tag{5.1}$$

subject to

$$[\omega.\phi(x_i)]\geq\rho-\xi_i\, and\, \xi_i\geq 0,\quad i=1,2,...,N,$$

where ω defines the optimal hyperplane, in which it is a normal vector to the hyperplane. The slack variables $\xi(\geqslant 0)$ are the error terms due to the misclassification. The specified apriori parameter υ in $(0,1]$, also known as the margin of the one-class SVM, corresponds to the probability of finding a new observation outside the separating boundary. Small values of υ result in less support vectors and smooth boundaries. However, large values of υ result in more support vectors and curvy, flexible boundaries. The optimal value of υ needs to be obtained to capture the complexity of the data and to avoid overfitting, respectively. The dual expression of 5.1 is

$$\underset{\lambda}{min}\frac{1}{2}\sum_{ij}\lambda_i\lambda_j K(x_i,x_j) \tag{5.2}$$

subject to

$$0\leq\lambda_i\leq\frac{1}{\upsilon N},\ i=1,2,...,N\ with\ \sum_i\lambda_i=1.$$

If ω and ρ solve this problem to find an optimal hyperplane, then the decision function would

be formed based on them as

$$f(x) = sgn[\omega.\phi(x) - \rho] = sgn[\sum_i \lambda_i K(x_i, x) - \rho]. \tag{5.3}$$

This function will be positive for most of the samples x_i in the training set based on a given assessment confidence. In general, this algorithm trains a hyperplane characterized by the pair (ω, ρ) which has maximal distance from the origin in the feature space H and separates all the data points from the origin. In original space, this algorithm trains a close boundary around the training samples so that, new samples which lay inside the boundary would be considered from the same population of training samples. If they lay outside, they would be considered abnormal with a given confidence in the assessment. This algorithm requires the choice of a kernel and a scalar parameter to define the boundary. Radial basis function (RBF) $K(x_i, x_j) = exp(-\|x_i - x_j\|^2/(2\sigma^2))$, where σ is a tuning parameter) [2] is usually selected and it is chosen in this research as well.

5.4 Validation Results and Discussions

5.4.1 Classification and Optimization Procedure

The experimental models, which are used in the proposed algorithm for performing classification and optimization, are models M, P, and C corresponding to data classes of type C_1, C_2, and C_3, respectively. The range of PD discharge magnitudes of these three single sources are more similar to each other and different from model F, therefore combinations which are composed of first three models generate multi-source PRPD patterns with more overlap levels. These partially overlapped patterns are more suitable to show performance potency and traceability of the proposed algorithm. However, model F (floating electrode) could be used in single-source PD classification block. The datasets, which are generated from experimental setups and used in this research, include a total of 300 data points for each of the 3 different single-source and each of the 4 different multi-source PD classes. For single-source defects, the data points form a dataset matrix $\mathbf{X}_{NJ\times 5L}$ whose dimension

is 900×500 (*i.e.* $L = 100$, $J = 300$, $N = 3$), and for multi-source of defects, we obtain a dataset matrix $\mathbf{Z}_{N'J\times 5L}$ with $L = 100$, $J = 300$, and $N' = 4$. Application of the PCA algorithm on matrix $\mathbf{X}$ results in a dimension reduction from 500 in the original space to 7 in the new informative space. The new dimension $k = 7$ is the appropriate dimensionality which corresponds to the intrinsic dimensionality of the data determined by the maximum likelihood estimation [2]. Application of PCA on single source data matrix $\mathbf{X}$ results in a mapping matrix $\mathbf{U}_{5L\times k}$ and the projected single source data, which forms dataset matrix $\mathbf{Y}_{5L\times k}$. The $\mathbf{U}$ matrix is then used to project multi-source samples on the same new low dimensional space by a simple multiplication (*i.e.* $\mathbf{T}_{4J\times k} = \mathbf{ZU}$).

In the proposed classification system, both linear and nonlinear[1] versions of LR could be employed separately. However, highly satisfying performance results of the algorithm with linear LR model are presented and analytical equations for linear LR models have been explained in this part. Regression coefficients of the first linear LR model are optimized by the application of LR model on the single source data samples (matrix $\mathbf{Y}$). The optimized coefficients have been recorded in matrix β_1. This helps to form analytical formula for logit stochastic models and posterior probabilities of all single source classes on the basis of all new low dimensional space variables. The regression function could be derived using the second LR model, which is applied on the posterior probabilities of single-source PDs. The coefficients of this regression function are optimized upon the application of multi-source samples available in the training set then those coefficients are recorded in matrix β_2. As mentioned before, only a limited number of multi-source samples in training part are enough for optimization of second LR model with only a few variables which would be equal to the number of available single source classes in training set. Using matrices β_1 and β_2, the analytical equations for logit stochastic models could be written as

[1]Nonlinear LR is more complex than linear LR and it could be employed to deal with highly nonlinear data samples

$$
\begin{aligned}
V = log\frac{P_i(C_1|X=x)}{P_i(C_3|X=x)} &= -1.4727 - 0.2579x_1 - 0.4866x_2 + 0.8487x_3 - \\
&\quad 0.0989x_4 - 0.0045x_5 + 0.0191x_6 - 0.0322x_7 \\
S = log\frac{P_i(C_2|X=x)}{P_i(C_3|X=x)} &= -2.7162 - 0.1687x_1 + 0.2350x_2 - 1.3992x_3 - \\
&\quad 0.8056x_4 - 0.5848x_5 - 0.1669x_6 - 0.0412x_7
\end{aligned} \tag{5.4}
$$

Posterior probability of a multi-source sample belonging to different single source classes of C_1-C_3 are derived using $Posterior_1 = (P_{C_1}, P_{C_2}, P_{C_3})$ with [6]

$$
\begin{aligned}
P_{C_1} = P_i(C_1|X=x) &= \frac{exp(V)}{1+exp(V)+exp(S)}, \\
P_{C_2} = P_i(C_2|X=x) &= \frac{exp(S)}{1+exp(V)+exp(S)}, \\
P_{C_3} = P_i(C_3|X=x) &= \frac{1}{1+exp(V)+exp(S)}.
\end{aligned} \tag{5.5}
$$

Also, the posterior probability of the samples belonging to multiple source classes are derived using the second set of logit stochastic models given by

$$
\begin{aligned}
Z' = log\frac{P_r(C_{1\&2}|P=P_i^*)}{P_r(C_{1\&2\&3}|P=P_i^*)} &= 19.8769 - 29.0721P_{C_1}^* - 41.2888P_{C_2}^* + 7.9180P_{C_3}^*, \\
S' = log\frac{P_r(C_{1\&3}|P=P_i^*)}{P_r(C_{1\&2\&3}|P=P_i^*)} &= 2.2570 - 1.3767P_{C_1}^* - 5.9847P_{C_2}^* + 19.5209P_{C_3}^*, \\
V' = log\frac{P_r(C_{2\&3}|P=P_i^*)}{P_r(C_{1\&2\&3}|P=P_i^*)} &= 3.0636 - 4.0931P_{C_1}^* - 10.5777P_{C_2}^* - 16.8282P_{C_3}^*
\end{aligned} \tag{5.6}
$$

where $(P_{C_1}^*, P_{C_2}^*, P_{C_3}^*)$ are obtained after implementing a PCA on $(P_{C_1}, P_{C_2}, P_{C_3})$ to make them uncorrelated and suitable for fitting the second LR model. The functional forms of posterior probabilities of multi-source classes for an unknown multisource sample would be equal to $Posterior_2 =$

$(P_{C_{1\&2}}, P_{C_{1\&3}}, P_{C_{2\&3}}, P_{C_{1\&2\&3}})$ where

$$\begin{aligned}
P_{C_{1\&2}} &= P_r(C_{1\&2}|P = P_i^*) = \frac{exp(Z')}{1 + exp(Z') + exp(S') + exp(V')}, \\
P_{C_{1\&3}} &= P_r(C_{1\&3}|P = P_i^*) = \frac{exp(S')}{1 + exp(Z') + exp(S') + exp(V')}, \\
P_{C_{2\&3}} &= P_r(C_{2\&3}|P = P_i^*) = \frac{exp(V')}{1 + exp(Z') + exp(S') + exp(V')}, \\
P_{C_{1\&2\&3}} &= P_r(C_{1\&2\&3}|P = P_i^*) = \frac{1}{1 + exp(Z') + exp(S') + exp(V')}.
\end{aligned} \tag{5.7}$$

To test the proposed algorithm and classify a new unknown multi-source sample, at first it is required to project it to the same space of single source training set using matrix **U**, and then pass it through the one-class SVM. If this sample doesnt belong to one of the single source classes then it should be entered into the traceable analytical equations derived in the training part. To summarize, entering the projected unknown sample into the first set of analytical equations (5.5) would generate posterior probabilities that the sample belongs to single source classes. Then passing these probabilities through a PCA algorithm and entering the resulting values into the analytical formulas of the second LR model (5.7) would estimate posterior probabilities that this sample belongs to different multi-source classes. At the end of the algorithm, class labeling would be performed by the Bayes classification rule which assigns the new sample to one of the multi-source classes $C_{1\&2}$, $C_{1\&3}$, $C_{2\&3}$, and $C_{1\&2\&3}$ corresponding to the one with the highest posterior probability. As a simple test, classification of 400 multi-source test samples is performed using analytical equations(5.5) and (5.7). This test subset contains 100 samples from each of the classes of $C_{1\&2}$, $C_{1\&3}$, $C_{2\&3}$, and $C_{1\&2\&3}$ $(400 = 100 + 100 + 100 + 100)$. Classification rate for this subset was 99.0%. The test confusion matrix (CM) evaluated for this test subset is

$$CM = \begin{bmatrix} 100 & 0 & 0 & 0 \\ 2 & 97 & 0 & 1 \\ 0 & 1 & 99 & 0 \\ 0 & 0 & 0 & 100 \end{bmatrix}$$

5.4.2 Performance Evaluation of the Algorithm

To perform a comprehensive evaluation of the proposed algorithm, the dimension-reduced, single-source dataset $\mathbf{Y}$ and multi-source dataset $\mathbf{T}$ are fed into the algorithm for both training and testing purposes. For efficiency verification, performance of the algorithm is evaluated based on the classification error rate. The algorithm at first is trained using training samples and then, it is tested on multi-source samples. The misclassified rate of test samples is considered as an estimate of the error rate. The single source dataset $\mathbf{Y}$ is used to optimize first LR model coefficients. To properly evaluate the prediction ability of the algorithm (including both linear and nonlinear LR) and also to tune the kernel parameters of LR models (in nonlinear cases), a $10-$fold cross validation (Rotation method [45]) is applied on dataset $\mathbf{T}$.

The $10-$fold cross validation algorithm has been selected over leave-one-out or holdout methods because of its less computational complexity, higher efficiency, and better performance on the PD subset [9, 45]. This method divides the multi-source dataset matrix $\mathbf{T}$ into 10 subsets of equal size, uses 9 subsets in training for the second LR model coefficients optimization and one for testing. This procedure is repeated 10 times until all training samples have been used for the training (second LR coefficients optimization) and exactly once for testing. Finally, the average error rates are reported to evaluate the accuracy of the classification performed by the proposed algorithm. This classification accuracy rates would be reported for algorithms with linear LR models. If one or both of the LR models in the algorithm is nonlinear, the value of kernel parameters for both or either of LR models needs to be tuned. The tuning process is performed by repeating the cross validation procedure until the average error rate reaches its minimum. Then optimum values of

kernel parameters and this minimum value of classification error rate will be reported. Finally, for the future prediction of unknown multi-source PDs using PRPD patterns, the algorithm will be trained using $\mathbf{Y}$ and the entire $\mathbf{T}$ matrix for the coefficients optimization of the first and second LR models, respectively.

Data samples for all classes of PD sources are captured under 2 levels of voltage equal to 20% and 50% higher than the inception voltage and two different noise levels. The classification accuracy rate evaluated for the proposed algorithm with linear LR models is given in Table . This table presents the classification success rate of our proposed algorithm with two types of dataset which are generated using an existing feature generation approach (3 univariate distributions of peak discharge, average discharge, and discharge rate) and the new feature generation approach which includes four highest quantiles of 200-quantiles plus peak discharge. These rates are presented for each individual multi-source classes of PD and the overall multi-source classes. As it is shown in Table 5.1, the proposed algorithm integrated with the new approach of feature generation shows a high accuracy rate of classification equal to 97.2%. This is a considerable accuracy rate for classification of a multi-source PD associated with PRPD patterns that are partially overlapped. Application of the proposed algorithm on the dataset generated by new feature generation algorithm outperforms its application on dataset made by the existing feature generation approach.

5.4.3 Performance Analysis of the Algorithm

The purpose of this part of my thesis is the identification of multi-source PDs which may simultaneously occur in HV apparatus. Identification of this type of PD sources is more practical and at the same time more complicated than identification of single sources of PD. To investigate this issue, one trivial approach, like the algorithms explained in Chapter 4 for single PD source identification, is to treat multi-source classes like single sources and generate a dataset composed of all classes. In this case, combination of matrices $\mathbf{X}$ and $\mathbf{Z}$ would form a seven-class dataset. After preprocessing, a traditional classification method has to be trained on the entire dataset and then tested. Despite

Table 5.1: Classification rate of proposed algorithm on data output of two different feature generation approaches. **Existing feature generation approach**. 3 univariate distributions of peak discharge, average discharge, and discharge rate; **new feature generation approach:** Four highest quantiles of 200-quantiles plus peak discharge.

	Class $C_{1\&2}$ %	Class $C_{1\&3}$ %	Class $C_{2\&3}$ %	Class $C_{1\&2\&3}$ %	Overall %
Existing Feature generation approach	88.3	53.0	99.7	75.7	79.2
New feature generation approach	**89.7**	**99.7**	**100.0**	**99.3**	**97.2**

the simplicity of this approach, data samples of multisources of PD would sometimes end up to partially overlap with single source data. This would show up in the form of data points belonging to multi-source classes which have been scattered in data space and mix up with data points of single source classes [31]. This issue becomes more severe as the number of classes increases and this is because, during the application of feature extraction, samples from classes with more similarity would be scattered with more overlapping clusters in the low dimensional feature space. This causes single source data and multi-source data from different classes which are more similar to get closer and overlap more. All these challenges make it more complex to perform PRPD based classification on multi-source samples with a satisfactory performance.

However, our proposed algorithm tackles these issues and considerably improves classification performance of the multi-source samples. In order to fulfil this we first separate single source and multi-source dataset and apply PCA only on the single-source dataset. Multi-source dataset for the test procedure of the algorithm are then projected onto the same space given by PCA. This separation of single source and multi-source classes decreases the overlapping level of classes so they become relatively separated in the low dimensional space [2].

In general, the proposed algorithm performs classification based on the probability estimation.

Probability of multi-source samples would be estimated with respect to all individual single-source classes available in training set (output of first LR model). However, the values of these probabilities, in general, will not imply which one of the individual sources contributed in making a multi-source sample. This happens sometimes because geometric configuration of the multi-source data points might be scattered close to the single source classes which may have not contributed to the generation of those samples. Relying just on the estimated values of these probabilities for classifying multi-source samples would be contradictory to their true constitutive single sources. This fact will causes ambiguity which leads to difficulties in classification. To overcome this issue, application of another regression function on these derived probabilities would be extremely beneficial. This regression function (second LR model) is aimed to truly model the relationship between the first step LR probabilities to exactly identify the constitutive single-sources which contribute to the formation of a multi-source sample. Handling these issues by the proposed algorithm implies its significance for the successful classification of multi-source PDs.

5.4.4 Risk Assesment Based on Probabilistic Interpretation

Table 5.1 presents the classification accuracy rates. These accuracy rates are reported based on the class labels which are assigned to the unknown multi-source test samples. The proposed algorithm not only shows high classification accuracy rates, but it also calculates the posterior probabilities of multi-source classes. This is one of the advantages of this algorithm which is accomplished due to the application of LR models. To demonstrate the posterior probabilities which are calculated by the proposed algorithm, 16 classified data samples were randomly selected. The probabilistic classification results for these samples are shown in Table 5.2. Each of these 16 samples leads to a different posterior probability that shows its "degree of membership" to a specific multi-source class. In this table, sample three, which is originally from class $C_{1\&2}$, is determined to belong to classes $C_{1\&2}$, $C_{1\&3}$, $C_{2\&3}$, and $C_{1\&2\&3}$ with probabilities of 82.2%, 10.4%, 0.2%, 7.2%, respectively. This probabilistic interpretation would allow safer decision making and has the following advantages:

Table 5.2: Classification rates for 16 test samples using the proposed algorithm

Data Point	**True Class**	Class $C_{1\&2}$ %	Class $C_{1\&3}$ %	Class $C_{2\&3}$ %	Class $C_{1\&2\&3}$ %
1	$C_{1\&2}$	**97.6**	0.0	0.0	2.4
2	$C_{1\&2}$	**99.8**	0.0	0.1	1.8
3	$C_{1\&2}$	**82.2**	10.4	0.2	7.2
4	$C_{1\&2}$	**72.4**	20.5	0.3	6.8
5	$C_{1\&3}$	27.4	**71.6**	0.3	0.7
6	$C_{1\&3}$	31.8	**67.0**	0.3	0.9
7	$C_{1\&3}$	12.6	**87.2**	0.2	0.0
8	$C_{1\&3}$	7.9	**92.0**	0.1	0.0
9	$C_{2\&3}$	3.9	1.9	**94.1**	0.1
10	$C_{2\&3}$	8.8	0.5	**90.5**	0.2
11	$C_{2\&3}$	2.6	0.2	**97.2**	0.0
12	$C_{2\&3}$	4.8	1.2	**94.0**	0.0
13	$C_{1\&2\&3}$	7.2	0.0	0.0	**92.8**
14	$C_{1\&2\&3}$	0.0	0.0	0.0	**100.0**
15	$C_{1\&2\&3}$	40.8	0.2	0.0	**59.0**
16	$C_{1\&2\&3}$	23.4	0.0	0.9	**75.7**

1. Rejecting a sample from classification by setting a threshold based on an acceptable "degree of membership" for a specific class or overall classes.

2. Considering the risk associated with different sources of PD in HV apparatus.

3. Referring a marginal classification to an expert operator.

4. Similarity level of a sample to the classes of data would be possible. This is mostly important for classification of a new unknown multi-source PRPD sample in the future.

Notably, for advantages $1-3$, the threshold for different classes would be defined based on the risk imposed by a specific multi-source PD for the safe operation of HV apparatus under test.

5.5 Summary

In this chapter of my thesis, a new algorithm has been developed for the identification of multi-source PDs which may simultaneously occur in HV insulation. Multi-source PDs sometimes result in partially overlapped patterns. This increases difficulties of identification of these types of patterns. To tackle this issue and conduct probabilistic identification of such sources, this chapter presented a comprehensive algorithm mainly constructed based on a novel two-step logistic regression model.

In order to successfully identify multi-source samples, PCA is applied on a database that is composed of single-source PDs. Then, multi-source samples are mapped on a low dimensional space derived by PCA algorithm. To differentiate multi-source test samples from single-source classes, one-class KSVM has been adopted on the data in that low dimensional space. Classification based on the probability estimation following two rounds of LR model is then executed. Initially, probability of multi-source samples are estimated using all individual single-source classes that are available in training set. A regression function is then employed to truly model the relationship between the first step LR probabilities to exactly identify a multi-source PRPD pattern. To evaluate the performance of the proposed algorithm for classification of the multi-source patterns, PD measurements on a number of multi-source setups have been conducted to simulate common defects of GIS

in small-scale laboratory test cells with realistic SF_6 gas condition. All datasets are made using fingerprints that are extracted by a novel approach from recorded PRPD patterns. Performance evaluation shows that the method works very well in identifying multi-source PRPD patterns based on their probabilistic relationship with individual single-source PRPD patterns. Further to its successful performance, this algorithm presents some sets of analytical equations which help to develop classification procedure in a simple processing task.

There are a few studies pertinent to the classification of multi-source PRPD that are mainly based on analyzing the PD pulse waveforms to separate individual concurrent pulse sources from the multi-source ones. Our proposed method has several advantages over the so-called separation/classification methods. The separation of individual pulse sources is possible due to the assumption of existing a correlation between the nature of PD sources and their generated pulse waveforms. This correlation is essential to properly distinguish different pulse waveforms originated from different sources. However, correlation assumption might not be completely true as PD pulse shape depends on the PD location in insulation and measurement system which is used. Moreover, application of separation/classification technique on all individual pulses is time consuming and dealing with noise-shape signals is also a challenge. Our proposed approach does not need to do separation of different pulse sources to identify them and it is based on PRPD pattern that is still a commonly used tool in the power industry.

Another important aspect of our proposed method is its capability of calculating a "degree of membership" to each multi-source class. This helps to take the risk associated with different PD sources that are activated in HV apparatus into account. The results of this research can be used to design a solid basis for an automated multi-source classification system and facilitate multi-source identification in early stages that would enhance the safety of HV apparatus, such as transformers, electric machines, cables, and Gas-Insulated Switchgear (GIS).

Chapter 6

Conclusions and Future Work

6.1 Conclusions

In this thesis, design, development, and testing of a comprehensive and automated classification system for single and multiple simultaneously activated PD sources identification based on the relationship between the variation of PRPD patterns and the source of PD was proposed. This automated classification system consists of two main parts; feature generation methods, and feature extraction and classifier algorithms that are implemented for the recognition of phase resolved partial discharge patterns.

To test the designed automated classification system on single PD source identification, different artificial laboratory test setups were designed and built to model different PD sources in a variety of insulation media and conditions. The first set includes test cells which are designed to model common sources of PD in air, oil, and SF_6. These test cells include 3 sources of partial discharge in SF_6 (floating electrode, moving particles, and fixed protrusion), 2 sources of PD in transformer oil (free particle and needle electrode), and corona in air. The second set is designed to test the classification system on the identification of partial discharge sources in oil-immersed insulation. These sources of partial discharge are simulated to model common PD sources in HV transformer

which include, bubble wrap to model small air bubbles, needle electrode in oil to simulate corona discharge, and floating metal particles (shavings) in oil. The third set is designed to show the performance of proposed PD classification system on the power transformer cellulose insulation samples under both electrical and thermal stresses. This capability enables online monitoring of high voltage cellulose insulation more accurately and efficiently which helps to prevent most transformer failures.

Using the proposed system, probabilistic interpretation of an unknown PRPD pattern that is to be classified was presented using some of the known classifier algorithms. These classifier algorithms, including Fuzzy classifiers (FSVM, FkNN) and Bayesian, are able to show a high accuracy rate of classification further to providing a knowledge of the "degree of membership" of a test sample to a class of data. This could be more beneficial rather than a class label assignment. Such knowledge enables probabilistic interpretation of an unknown PRPD pattern that is being classified. Overall, these classification results and availability of posterior probability show prosperous performance in this area of studies and to some extent indicate the promising possibility of online and offline automatic classification of single PD sources in HV apparatus. Besides, the results of automated PD classification system on both oil-immersed insulation and thermally-degraded cellulose-oil insulation show prosperous performance in these fields as well.

In online condition assessment monitoring of high voltage (HV) insulators, beside single PD source identification, it is sometimes required to identify multiple, simultaneously activated partial discharge (PD) sources that happen in the insulation of the HV apparatus. Multi-source PDs sometimes result in partially overlapped patterns. This increases difficulties of identification of these types of patterns.

To tackle this issue and conduct probabilistic identification of such sources and to further enhance the proposed classification system, a novel algorithm for identification of the PRPD pattern that is a mix of multiple, simultaneous PD sources has been developed and appended to the proposed system. In fact, we developed a novel algorithm to identify multiple, simultaneously activated PD

sources using PRPD patterns that are widely used in power industry and are easier to analyze compared to time-consuming PD pulse waveforms analysis which has been used by other researchers. The multi-source PRPD pattern classification, developed using training and test databases that are generated from fingerprints of single-source PD patterns and probabilistic interpretation, is performed following a novel two-step Logistic Regression (LR) algorithm [6]. This two-step LR algorithm is trained on the database derived from single-source patterns. The algorithm is then tested on samples that are generated with multi-source PRPD patterns.

To evaluate the performance of our proposed algorithm for classification of multi-source patterns, PD measurements on a number of multi-source models are conducted. These artificial models are built to simulate common defects of GIS in small-scale laboratory test cells with realistic SF_6 gas condition.

Several studies have been performed by other researchers for classification of multi-source PDs. These are mainly based on analyzing the PD pulses to firstly separate individual pulse sources, then classifying them into their belonged classes. The separation of individual pulse sources, in these researches, are mainly conducted based on the correlation which is assumed to exist between the nature of PD sources and their generated pulse waveforms. The existence of this correlation is the base of this classification method in the area of multi-source PD identification and it is essential to conduct proper discrimination between different sources using their pulse waveforms. However, this correlation assumption might not be always true as PD pulse shape also depends on other parameters, such as PD location in insulation and specification of components which are used in measurement system. Another drawback of separation/classification technique is that it requires to analyze all individual pulses which is a time consuming procedure. Also dealing with noise shape signals is another challenge in this technique. Our proposed method in this research has several advantages compare to this separation/classification method. The proposed method does not require to do separation of different pulse sources to be able to identify them. This method works using PRPD pattern that is still a commonly used tool in the electric power

industry. Besides the prosperous performance of this proposed algorithm, it presents some sets of analytical equations which help to conduct classification procedure of multi-source PDs with an easy processing. Availability of these analytical equations would be significant achievements for identification of unknown data arising from HV insulation system in future. Further, an important problem in PD source identification is to assign "degrees of membership" to multi-source PRPD patterns besides assigning a class label to them. This enables to pwerform probabilistic classification of a new, unknown multi-source sample. The availability of this degree of membership for future PRPD samples would allow safer decision-making by taking the risk of various sources of PDs in HV insulation system into account. The results of this work show capability to design a solid basis for an automated multi-source classification system and facilitate PD source identification in early stages. This could help to increase safety of HV apparatus, such as transformers, electric machines, cables, and GIS. This achievement assists in detection and identification of multiple PD sources in early stages which ends up to prevent costly failures of electrical equipment.

In summary, the results of automated classification system on PD sources identification based on different feature extraction and classification algorithms were presented. The results show that the proposed classification system is well able to successfully identify the single and multiple sources of partial discharge in different insulation media under different working temperatures. Availability of this classification system enables continuous 24/7 monitoring of equipments and helps to identify PD sources in early stages which leads to safe operation of HV apparatus. Also providing the probabilistic interpretation based on the risk associated with different PD sources, a marginal classified PRPD sample by the proposed classification system will be referred to an expert operator to do a visual inspection and make a proper decision; otherwise, decision will be made by the classification system.

To further enhance the classification system, a thorough dataset is required that must encompass all types of PD sources. This is to avoid rejection and reduce misclassification rate for PDs from all possible single source and multi-sources and construct more powerful classification system.

The approach used in this research has the ability at append PRPD data generated from other types of single and multiple PD sources to construct more powerful classification system.

6.2 Main Contributions

In this PhD thesis, an automated classification system for both single and multiple PD source identification was proposed. This system investigated the relationship between the sources of PD and the variation of specific features of PRPD patterns that were acquired from those sources of PD. Based on this correlation, this system was trained and optimized to predict class label and probability of a sample belonging to the known sources of PD. Proposed system consists of feature generation, feature extraction and classifier algorithms that would be implemented on the PRPD patterns for identification of the source of partial discharge.

This PhD thesis contributes in the areas of:

- Classification of single PD sources using 12 high performance, applicable methods on PRPD pattern data for dimensionality reduction (including the traditional statistical operators) which are chosen exploring almost all available well-developed feature extraction techniques, as well as 10 well-known algorithms for classification have been explored. The classification success rate of their application on the PD patterns of the discharge activities in different insulation media including air, oil, SF_6 has been evaluated.

- Some of the classifier algorithms developed in this work, such as fuzzy classifiers, are not only capable to show high classification accuracy rate, but they also calculate the "degree of membership" of a sample to a class of data. This enables probabilistic interpretation of a new PRPD pattern that is being classified.

- The availability of this degree of membership for future PRPD samples would allow safer decision making based on the risk associated with different sources of PD in HV apparatus.

- Test sets are designed to study PRPD patterns and show the performance of proposed classification system on identification of single partial discharge sources in oil-immersed insulation under electrical stress and the power transformer cellulose insulation samples under both electrical and thermal stresses. This capability enables online monitoring of high voltage cellulose insulation more accurately and efficiently which helps to prevent most transformer failures.

- To generate a dataset, two old approaches which have been used in past [4] are modified in this work to considerably increase their discriminatory power.

- A new approach for feature generation is proposed with strong discrimination power to differentiate between PRPD patterns of different sources. The efficiency of this approach will be considerably perceived dealing with multiple simultaneously activated PD sources in HV insulation.

- Classification of the PRPD pattern that is a mix of multiple, simultaneous PD sources was performed. To do so, we developed a novel algorithm to identify multiple, simultaneously activated PD sources using PRPD patterns that are widely used in power industry and are easier to analyze compared to PD pulse waveforms analysis. The multi-source PRPD pattern classification is developed using training and test databases that are generated from fingerprints of single-source PD patterns and probabilistic interpretation is performed following a novel two-step Logistic Regression (LR) algorithm [6].

In summary, proposed classification system has several advantages in HV insulation monitoring which are:

* Classification of PRPD pattern with high accuracy rate.

* Probabilistic interpretation based on the “Degree of Membership.”

* Risk assessment of different sources of PD in HV apparatus.

* Investigation of similarity between different sources of PD.
* Referring of a marginal classification to an expert operator.
* High classification rate of multiple, simultaneous PD sources.
* Enabling continuous 24/7 monitoring.
* Considering the effects of different parameters such as the increase trend of work temperature on generated PRPD patterns.

6.3 Future Work

Proposed classification system described in this thesis is tested based on PD measurements recorded in a controlled lab environment. The next steps for continuing this work involve testing the proposed system on data which is acquired from real HV apparatus working online or offline in the electric power systems.

Capability of conducting PD risk assessment based on probabilistic interpretation of PD samples is one of the advantages of this proposed classification system; however, different types of fault will have different costs and consequences which are also dependent on the HV equipment which are occurring in, therefore taking all these factors into account via a cost model might be useful to define the appropriate threshold level for accepting the classification of PD samples. This would be beneficial method for an automated and online monitoring system which worth to be considered as a future work in this area of research.

Further, classification could be continued analyzing the PD waveforms of proposed PD source setups and exploring the application of high performance algorithms of machine learning for classification of PD pulse sources. Some efforts need to be taken to reduce the costs of this type of analysis method (discussed in the previous chapter) and increase its reliability and success rate. This can be done using Wavelet Transform and Orthogonal Polynomial Expansions which will be

exerted on the entire PD waveforms to employ their important and well-discriminative features. Finally, accuracy estimation rate of the application of this method on different PD pulse waveform arose from designed insulation setups in this research could be evaluated and compared to the proposed technique in this thesis using PRPD patterns.

Another important task which could be continued in future, would be based on the use of orthogonal polynomial expansion (especially, Hermite polynomial, and Laguerre polynomial) and their properties which helps on performing system identification.

These suggestions for future work are explained in details in the following subsections.

6.3.1 PD Waveform General Framework

A set of discharges from a specific PD source approximately have similar waveforms. However, these waveforms are assumed to be different than the waveforms from another PD sources. In the proposed classification system in this thesis, the apparent charge amplitude (plotted in PRPD patterns) has been used and the detection system has been calibrated based on this charge amplitude. All the phase resolved PD patterns and time resolve PD patterns generated only based on the PD charge amplitude. However, PD signals arise from different sources of PD, have other specific features which might be of help in achieving better classification results. Using PD waveforms seems to be promising so it is worth to be considered[1]. One comprehensive method could be the application of algorithms such as Wavelet Transforms or Orthogonal Polynomial expansions which will be applied on the entire PD waveforms. These algorithms take the entire waveform as the input and based on its shape, they would generate informative data at the output. This data will be used as the new input for the classification algorithms. Using this method, multiple, simultaneous PD source classification could also be performed. However, as mentioned in the previous chapter, there are some challenges associated with this PD waveform analysis technique which need to be considered and tackled. If these challenges are resolved, it is speculated that, prosperous multi-

[1]This seems to be promising if the drawbacks of this method could be resolved.

ple, simultaneous PD sources classification could be possible using PD waveforms originated from different types of PD sources. At the end, the accuracy estimation rate of the application of this method on different insulation media including air, oil, SF_6 could be evaluated and compared to the available classification system results. Another important achievement of the use of Orthogonal Polynomial Expansion could be the potential of performing system identification. The potential of system identification as another future work suggestion will be explained in the last subsection of this chapter.

6.3.1.1 Wavelet Transform

A powerful method which can be implemented for analyzing a PD signal and extracting some specific features, is Wavelet Transform [5, 25, 30, 38]. An efficient, equivalent way to implement discrete wavelet transform is through the use of a filtering algorithm. The output of this algorithm is wavelet coefficients which are derived after passing the signal through a set of low pass and high pass filters. The output of low pass filter then again is filtered for constructing the next level. The output of low pass and high pass filters at each scale builds approximation coefficients (CA_s) and detail coefficients (CD_s), respectively. These coefficients contain the signal's characteristics which can be utilized to rebuild the signal in the reconstruction process. Two methods could be implemented on these coefficients in the M levels of decomposition to make feature vectors that will characterize PDs. One approach is the implementation of the first four orders of moments (mean, variance, Skewness, and Kurtosis) on the probability distribution of coefficients in each M decomposition levels. Therefore, a features vector for each PD signal would be formed with a dimension equal to $4M$. These vectors could be used as the input of the classification stage. The second approach is the application of linear and nonlinear feature extraction algorithms (*e.g.* PCA and FDA) on the histograms of these coefficients. Using the histogram of the coefficients at each level (with N bins) forms a vector of $9N$ elements for each PD signal. Subsequently, different feature extraction algorithms could be utilized for dimension reduction of the data. This would

make the feature vector being more proper for conducting an accurate classification. Multiple, simultaneous PD sources classification could also be possible using PD waveforms and applying Wavelet transform on them.

6.3.1.2 Polynomial Expansions

As mentioned before, a set of discharges from a specific PD source approximately have similar waveforms. Orthogonal polynomials are classes of polynomial functions defined over a specific range and obey an orthogonality relation. These polynomials have useful properties which make them powerful in dealing with mathematical and physical problems. Some common orthogonal polynomials are, Chebyshev polynomial, Gegenbauer polynomial, Hermite polynomial, Jacobi polynomial, Laguerre polynomial, and Legendre polynomial. Each one of these orthogonal polynomials has specific features which make them powerful in dealing with specific problems. Orthogonal polynomials in general work as a powerful algorithm and take the entire waveform which has a finite time as the input and decompose it into a series expansion. This approach would be very promising for classification of PD because of its capability for information compression. In the last few years, considerable attention has been given to the use of the orthogonal polynomials; however most of the applications have been on image processing rather than on mono-dimensional signals.

6.3.2 System Identification

Another important achievement of the use of orthogonal polynomial expansion could be the ability of performing system identification. This would be achieved by the important feature of Hermite polynomial and Laguerre polynomial related to the simple equation for convolution of each pair of their basis functions. This feature in fact enables us to analytically de-convolve the expanded PD signals by expanded impulse response of the system and get the actual PD signal which is happening in the exact location of the test cell. In this way, the actual PD waveform with no effect of the system on it could be detected. Besides, it will become possible to analyze the effect of the

system on the PD signal by comparing the input and output signals. It is worth mentioning the issue with the effect of the system on the PD pulse shape which was explained as a drawback of PD pulse waveform analysis method (separation/classification method) and that it can be addressed using the system identification. Another important achievement of performing system identification would be the possibility of running the classification on real unaffected PD pulses. This makes all the classification procedure independent of the system. In other words, data bank collected in a laboratory (or company) becomes operational in other laboratories (or companies).

References

[1] J. C. Fothergill, "Ageing, space charge and nanodielectrics: Ten things we don't know about dielectrics," in *2007 International Conference on Solid Dielectrics, ICSD*, 2007, pp. 1–10.

[2] K. K. S. Theodoridis, *Pattern Recognition*, 4th ed. Elsevier Inc, 2009.

[3] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: John Wiley & Sons, 2001.

[4] A. Krivda, "Automated recognition of partial discharges," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 5, pp. 796–821, 1995.

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

[6] C. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[7] H. Janani, N. D. Jacob, and B. Kordi, "Automated recognition of partial discharge in oil-immersed insulation," in *Electrical Insulation Conference (EIC), 2015 IEEE*, June 2015, pp. 467–470.

[8] H. Janani, N. Jacob, and B. Kordi, "Partial discharge pattern recognition for thermally degraded cellulose-oil insulation," in *CIGRE Canada*, Winnipeg, Manitoba, August 2015.

[9] H. Janani and B. Kordi, "Towards Automated Partial Discharge Source Classification," *IEEE Transactions on Power Delivery*, Submitted, 2016.

[10] H. Janani, M. J. Jozani, and B. Kordi, "Classification of Simultaneous Multiple Partial Discharge Sources Based on Probabilistic Interpretation Using a Two step Logistic Regression Algorithm," *IEEE Transactions on Dielectrics and Electrical Insulatoin*, Submitted, 2016.

[11] H. Janani, N. Jacob, M. J. Jozani, and B. Kordi, "Probabilistic Analysis of Partial Discharges in Thermally-degraded Power Transformer Cellulose-oil Insulation Using Pattern Recognition Techniques," *IEEE Transactions on Power Delivery*, Submitted, 2016.

[12] IEC 60270, "High-voltage test techniques - Partial discharge measurements," Tech. Rep., 2001.

[13] *Standard Test Method for Detection and Measurement of Partial Discharge (Corona) Pulses in Evaluation of Insulation Systems*, Std. ASTM D1868 Std., 2007.

[14] A. Krivda, "Recognition of discharges," Ph.D. dissertation, Delft University, 1995.

[15] A. N. Arman and A. T. Starr, "The measurement of discharges in dielectrics," *Journal of the Institution of Electrical Engineers*, vol. 79, no. 475, pp. 67–81, July 1936.

[16] D. M. Robinson, *Dielectric Phenomena in High Voltage Cables*. Chapman and Hall, 1936.

[17] R. E. James and B. T. Phung, "Development of computer-based measurements and their application to PD pattern analysis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 5, pp. 838–856, 1995.

[18] H.-G. Kranz and R. Krump, "Partial discharge diagnosis using statistical optimization on a PC-based system," *IEEE Transactions on Electrical Insulation*, vol. 27, no. 1, 1992.

[19] M. M. A. Salama and R. Bartnikas, "Determination of neural-network topology for partial discharge pulse pattern recognition," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 446–456, 2002.

[20] L. A. Petrov, P. L. Lewin, and T. Czaszejko, "On the applicability of nonlinear time series methods for partial discharge analysis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 21, no. 1, pp. 284–293, February 2014.

[21] C. Cachin and H. J. Wiesmann, "PD recognition with knowledge-based preprocessing and neural networks," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 4, pp. 578–589, 1995.

[22] E. Gulski and F. H. Kreuger, "Computer-aided recognition of discharge sources," *IEEE transactions on electrical insulation*, vol. 27, no. 1, pp. 82–92, 1992.

[23] Y. Han and Y. H. Song, "Using improved self-organizing map for partial discharge diagnosis of large turbogenerators," *IEEE Transactions on Energy Conversion*, vol. 18, no. 3, pp. 392–399, 2003.

[24] N. Hozumi, T. Okamoto, and T. Imajo, "Discrimination of partial discharge patterns using a neural network," *IEEE Transactions on Electrical Insulation*, vol. 27, no. 3, 1992.

[25] E. M. Lalitha and L. Satish, "Wavelet analysis for classification of multi-source PD patterns," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 7, no. 1, pp. 40–47, 2000.

[26] A. Mazroua, M. Salama, and R. Bartnikas, "PD pattern recognition with neural networks using the multilayer perceptron technique," *IEEE Transactions on Electrical Insulation*, vol. 28, no. 6, 1993.

[27] L. Satish and W. Zaengl, "Artificial neural networks for recognition of 3-d partial discharge patterns," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 1, no. 2, 1994.

[28] M.-H. Wang, "Partial Discharge Pattern Recognition of Current Transformers Using an ENN," *IEEE Transactions on Power Delivery*, vol. 20, no. 3, pp. 1984–1990, 2005.

[29] A. Contin, A. Cavallini, G. C. Montanari, G. Pasini, and F. Puletti, "Digital detection and fuzzy classification of partial discharge signals," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 9, no. 3, pp. 335–348, 2002.

[30] A. Cavallini, G. C. Montanari, F. Puletti, and A. Contin, "A new methodology for the identification of PD in electrical apparatus: Properties and applications," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 12, no. 2, pp. 203–215, 2005.

[31] H. Ma, J. C. Chan, T. K. Saha, and C. Ekanayake, "Pattern recognition techniques and their applications for automatic classification of artificial partial discharge sources," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 20, no. 2, pp. 468–478, April 2013.

[32] R. M. Sharkawy, R. S. Mangoubi, T. K. Abdel-Galil, M. M. A. Salama, and R. Bartnikas, "SVM classification of contaminating particles in liquid dielectrics using higher order statistics of electrical and acoustic PD measurements," in *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 14, no. 3, 2007, pp. 669–678.

[33] L. Satish and B. Gururaj, "Use of hidden Markov models for partial discharge pattern classification," *IEEE Transactions on Electrical Insulation*, vol. 28, no. 2, 1993.

[34] A. Contin and G. Rabach, "PD analysis of rotating ac machines," *IEEE transactions on electrical insulation*, vol. 28, no. 6, pp. 1033–1042, 1993.

[35] T. Abdel-Galil, R. Sharkawy, M. Salama, and R. Bartnikas, "Partial discharge pulse pattern recognition using an inductive inference algorithm," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 12, no. 2, 2005.

[36] X. Peng, C. Zhou, D. Hepburn, M. D. Judd, and W. H. Siew, "Application of K-Means method to pattern recognition in on-line cable partial discharge monitoring," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 20, no. 3, pp. 754–761, 2013.

[37] T. Okamoto and T. Tanaka, "Novel partial discharge measurement computer-aided measuremnet systems," *IEEE Transactions on Electrical Insulation*, vol. EI-21, no. 6, pp. 1015–1019, Dec 1986.

[38] D. Evagorou, a. Kyprianou, P. Lewin, a. Stavrou, V. Efthymiou, a.C. Metaxas, and G. Georghiou, "Feature extraction of partial discharge signals using the wavelet packet transform and classification with a probabilistic neural network," *IET Science, Measurement & Technology*, vol. 4, no. 3, p. 177, 2010.

[39] K. Wang, R. Liao, L. Yang, J. Li, S. Grzybowski, and J. Hao, "Optimal features selected by nsga-ii for partial discharge pulses separation based on time-frequency representation and matrix decomposition," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 20, no. 3, pp. 825–838, June 2013.

[40] J. A. Ardila-Rey, J. M. Martínez-Tarifa, G. Robles, and M. Rojas-Moreno, "Partial discharge and noise separation by means of spectral-power clustering techniques," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 20, pp. 1436–1443, 2013.

[41] L. Hao, P. Lewin, J. a. Hunter, D. Swaffield, a. Contin, C. Walton, and M. Michel, "Discrimination of multiple PD sources using wavelet decomposition and principal component analysis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 18, no. 5, pp. 1702–1711, 2011.

[42] J. M. Martinez-Tarifa, J. A. Ardila-Rey, and G. Robles, "Automatic selection of frequency bands for the power ratios separation technique in partial discharge measurements: part i, fundamentals and noise rejection in simple test objects," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 22, no. 4, pp. 2284–2291, August 2015.

[43] K. Wang, J. Li, S. Zhang, R. Liao, F. Wu, L. Yang, J. Li, S. Grzybowski, and J. Yan, "A hybrid algorithm based on s transform and affinity propagation clustering for separation of two simultaneously artificial partial discharge sources," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 22, no. 2, pp. 1042–1060, April 2015.

[44] J. C. Chan, H. Ma, and T. K. Saha, "Time-frequency sparsity map on automatic partial discharge sources separation for power transformer condition assessment," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 22, no. 4, pp. 2271–2283, August 2015.

[45] A. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

[46] E. Gulski, "Digital analysis of partial discharges," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 5, 1995.

[47] L. J. P. Van Der Maaten, E. O. Postma, and H. J. Van Den Herik, "Dimensionality Reduction: A Comparative Review," *Journal of Machine Learning Research*, vol. 10, pp. 1–41, 2009.

[48] H. Hotelling, "Analysis of a complex of statistical variables into principal components." pp. 417–441, 1933.

[49] R. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[50] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem," pp. 1299–1319, 1998.

[51] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach." *Neural computation*, vol. 12, no. 10, pp. 2385–2404, 2000.

[52] T. F. Cox and M. A. A. Cox, *Multidimensional scaling*. London, UK: Chapman & Hall, 1994.

[53] D. K. Agrafiotis, "Stochastic proximity embedding," *Journal of Computational Chemistry*, vol. 24, no. 10, pp. 1215–1221, 2003.

[54] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction." *Science (New York, N.Y.)*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[55] G. E. Hinton and S. T. Roweis, "Stochastic neighbor embedding," *Advances in neural information processing systems*, pp. 833–840, 2002.

[56] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding." *Science (New York, N.Y.)*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[57] X. He and P. Niyogi, "Locality preserving projections," *Neural information processing systems*, vol. 16, p. 153, 2004.

[58] X. H. X. He, D. C. D. Cai, S. Y. S. Yan, and H.-J. Z. H.-J. Zhang, "Neighborhood preserving embedding," *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 2, 2005.

[59] M. Hearst, S. Dumais, E. Osman, J. Platt, and B. Scholkopf, "Support vector machines," pp. 18–28, 1998.

[60] C.-F. Lin and S.-D. Wang, "Fuzzy support vector machines." *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, vol. 13, no. 2, pp. 464–471, 2002.

[61] J. Keller, J.M. and Gray, M.R. and Givens, "A fuzzy K-nearest neighbor algorithm," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. SMC-15, no. 4, pp. 580–585, 1985.

[62] C. Bishop, *Neural Networks for Pattern Recognition*. Oxford Univ. Press, 1995, vol. 92.

[63] D. F. Specht, "Probabilistic neural networks," pp. 109–118, 1990.

[64] Y. Freund and R. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," *Computational learning theory*, vol. 55, no. 1, pp. 119–139, 1995.

[65] R. Batuwita and V. Palade, "Fsvm-cil: Fuzzy support vector machines for class imbalance learning," *IEEE Transactions on Fuzzy Systems*, vol. 18, no. 3, pp. 558–571, June 2010.

[66] D. Specht, "probabilistic neural networks," *Neural Networks*, vol. 3, no. 1, pp. 109–118, 1990.

[67] P. Karsmakers, K. Pelckmans, and J. Suykens, "Multi-class kernel logistic regression: a fixed-size implementation," in *Neural Networks, 2007. IJCNN 2007. International Joint Conference on*, Aug 2007, pp. 1756–1761.

[68] S. Rahayu, S. Purnami, and A. Embong, "Applying kernel logistic regression in data mining to classify credit risk," in *Information Technology, 2008. ITSim 2008. International Symposium on*, vol. 2, Aug 2008, pp. 1–6.

[69] Z. Wu, Q. Wang, A. Plaza, J. Li, L. Sun, and Z. Wei, "Real-time implementation of the sparse multinomial logistic regression for hyperspectral image classification on gpus," *Geoscience and Remote Sensing Letters, IEEE*, vol. 12, no. 7, pp. 1456–1460, July 2015.

[70] B. F. Hampton and R. J. Meats, "Diagnostic measurements at UHF in gas insulated substations," *IEE Proceedings-Generation, Transmission and Distribution*, vol. 135, no. 2, pp. 137–144, 1988.

[71] A. Reid, M. Judd, R. Fouracre, B. Stewart, and D. Hepburn, "Simultaneous measurement of partial discharges using IEC60270 and radio-frequency techniques," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 18, no. 2, pp. 444–455, 2011.

[72] D. Kopejtkova, T. Molony, S. Kobayashi, and I. M. Welch, "A twenty-five year review of experience with has-insulated substations," *CIGRE Paper*, vol. 23, p. 101, 1992.

[73] S. Xiao, P. Moore, and M. Judd, "Investigating the assessment of insulation integrity using radiometric partial discharge measurement," in *International Conference on Sustainable Power Generation and Supply, 2009. SUPERGEN '09.*, April 2009, pp. 1–7.

[74] R. Piccin, A. R. Mor, P. Morshuis, A. Girodet, and J. Smit, "Partial discharge analysis of gas insulated systems at high voltage AC and DC," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 22, no. 1, pp. 218–228, 2015.

[75] H. Saadat, *Power System Analysis*, 3rd ed. McGraw Hill, USA, 2002.

[76] X. Chen, A. Cavallini, and G. C. Montanari, "Statistical analysis and fuzzy logic identification of partial discharge in paper-oil insulation system," in *Properties and Applications of Dielectric Materials, 2009. ICPADM 2009. IEEE 9th International Conference on the*, July 2009, pp. 505–508.

[77] H. Shiota, H. Muto, H. Fujii, and N. Hosokawa, "Diagnosis for oil-immersed insulation using partial discharge due to bubbles in oil," in *Properties and Applications of Dielectric Materials, 2003. Proceedings of the 7th International Conference on*, vol. 3, June 2003, pp. 1120–1123 vol.3.

[78] X. Chen, A. Cavallini, and G. C. Montanari, "Improving high voltage transformer reliability through recognition of pd in paper/oil systems," in *High Voltage Engineering and Application, 2008. ICHVE 2008. International Conference on*, Nov 2008, pp. 544–548.

[79] N. D. Jacob, D. R. Oliver, S. S. Sherif, and B. Kordi, "Statistical texture analysis of morphological changes in pressboard insulation due to thermal aging and partial discharges," in *Electrical Insulation Conference (EIC), 2015 IEEE*, June 2015, pp. 610–613.

[80] R. Sarathi, I. P. M. Sheema, and J. Rajan, "Understanding surface discharge activity in copper sulphide diffused oil impregnated pressboard under ac voltages," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 21, no. 2, pp. 674–682, April 2014.

[81] P. M. Mitchinson, P. L. Lewin, B. D. Strawbridge, and P. Jarman, "Tracking and surface discharge at the oil 2013;pressboard interface," *IEEE Electrical Insulation Magazine*, vol. 26, no. 2, pp. 35–41, March 2010.

[82] B. Buerschaper, O. Kleboth-Lugova, and T. Leibfried, "The electrical strength of transformer oil in a transformerboard-oil system during moisture non-equilibrium," in *Electrical Insulation and Dielectric Phenomena, 2003. Annual Report. Conference on*, Oct 2003, pp. 269–272.

[83] D. Garcia, R. Villarroel, B. Garcia, and J. Burgos, "Effect of the thickness on the water mobility inside transformer cellulosic insulation," *IEEE Transactions on Power Delivery*, vol. PP, no. 99, pp. 1–1, 2015.

[84] Y. Cui, H. Ma, T. Saha, and C. Ekanayake, "Understanding moisture dynamics and its effect on the dielectric response of transformer insulation," *IEEE Transactions on Power Delivery*, vol. 30, no. 5, pp. 2195–2204, Oct 2015.

[85] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, a. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution." *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[86] I. I. S. 60270, "High voltage test techniques- partial discharge measurements," *International Electrotechnical Commission(IEC), Geneva, Switzerland*, vol. 3rd edition, 2000.

[87] B. Hampton and R. Meats, "Diagnostic measurements at UHF in gas insulated substations," *Generation, Transmission and Distribution, IEE Proceedings C*, vol. 135, no. 2, pp. 137–145, Mar 1988.

[88] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment," *SIAM J. Scientific Computing*, vol. 26, no. 1, pp. 313–338, 2004.

[89] R. Rivest, T. Cormen, C. Leiserson, and C. Stein, *Introduction to Algorithms.* Cambridge, MA: MIT Press, 2001.

[90] V. Kumar, A. Grama, A. Gupta, and G. Karypis, *Introduction to Parallel Computing.* Redwood City, CA: Benjamin Cummings, 1994.

[91] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectraltechniques for embedding and clustering," *Neural Information Processing System*, vol. 14, pp. 585–591, 2001.

[92] D. Donoho and C. Grimes, "Hessian eigenmaps: New locally linear embedding techniques for high-dimensional data," *Proc. Nat'l Academy of Sciences*, vol. 100, no. 10, pp. 5591–5596, 2003.

[93] T. Zhang, J. Yang, D. Zhao, and X. Ge, "Linear local tangentspace alignment and application to face recognition," *Neurocomputing*, vol. 70, pp. 1547–1553, 2007.

Appendix A

Dimension Reduction Techniques

A.1 Linear Techniques

A.1.1 Principal Component Analysis (PCA)

Principal Component Analysis [48] is one of the most well-known unsupervised methods for dimensionality reduction in pattern recognition. This technique performs dimensionality reduction by projection of data onto a space with lower dimensionality based on capturing as much of the variance in the data as possible [2]. PCA searches for linear mapping U that would maximize $U^T cov(X)U$, which $cov(X)$ is the covariance matrix of the data. The data mapping which captures more percentage of variance after projection of the data can be derived by solving the Eigen-equation

$$\underset{X-\bar{X}}{Cov}.u - \lambda u = 0 \tag{A.1}$$

under the constraint of $||u_j|| = 1$. The eigenvectors of the covariance matrix form the columns of the linear mapping matrix U. It is worth noting that the covariance matrix of the mapped dataset will be diagonal which means that PCA suppresses cross-dimensional co-activity of variables and this is another advantage of PCA.

A.1.2 Fisher Discriminant Analysis (FDA)

Fisher Discriminant Analysis (FDA) algorithm [74] is a supervised algorithm that attempts to find directions for data projection that are efficient for discrimination of different classes in the low-dimensional representation [49]. FDA works based on the concept that means of different classes after projection get separated as far as possible while data points around their means having a small variance in there classes. Considering between-class scatter(S_b) and within-class scatter (S_w) matrices, optimal weight vectors (w) can be derived when maximizing the Fisher criterion

$$FDR(w) = \frac{w^T S_b w}{w^T S_w w} \tag{A.2}$$

with respect to w. It is known that this maximization leads to the following Eigen-problem

$$S_b w = \lambda S_w w \tag{A.3}$$

where λs are the largest eigenvalues of $S_w^{-1} S_b$ and the eigenvectors w form the linear mapping matrix W.

A.2 Non-Linear Techniques

A.2.1 Kernel Principal Component Analysis (KPCA)

Kernel PCA (KPCA) [50] is a nonlinear form of Principal Component Analysis (PCA). In other word, KPCA is a kernelized version of PCA which performs nonlinear feature extraction based on the use of the kernel approach. KPCA algorithm starts with making an implicit mapping of data set onto a new space with higher dimensions typically via a nonlinear function $\phi(x)$. Then it performs standard PCA in the mapped space which is often very high dimensional. In order to do this, the kernel function is used to construct the matrix K with entries equal to $K(i,j) = k(x_i, x_j)$ for all N data points. As it is shown in [50], after doing centering operation of kernel matrix, the

d dominant eigenvectors of K can represent eigenvectors of the covariance matrix in that mapped space. The low dimensional data Y, based on the normalized eigenvectors of matrix $K(named, a_m)$ and kernel function can be calculated as

$$Y = (\sum_j a_1^j k(x_j, x), \sum_j a_2^j k(x_j, x), ..., \sum_j a_d^j k(x_j, x)) \tag{A.4}$$

KPCA is greatly dependent on the type of kernel function which is applied. The usual applicable kernel functions are linear, the Gaussian and Polynomial kernels. Linear kernel leads the KPCA becomes equal to the standard PCA.

A.2.2 Generalized Discriminant Analysis (GDA) or Kernel FDA

Generalized Discriminant Analysis (GDA) [51] is a nonlinear dimension reduction algorithm based on extension of FDA. GDA is a linear algebraic formula in the transformed space which could be achieved using kernel operators. In fact, GDA performs FDA on the data after the input space has been mapped into a high dimensional space that maximizes seperability of the classes. This algorithm, similar to KPCA, starts by mapping the data into a Hilbert space using a nonlinear mapping function ($\phi(x)$). It is shown in [51], with the help of kernel operator, nonlinear computation in the original space changes to linear computation in the mapped space. Subsequently, maximizing the below Rayleigh quotient is equal to maximizing the fisher criterion in the mapped space as

$$\lambda = \frac{\alpha^T K W K \alpha}{\alpha^T K K \alpha} \tag{A.5}$$

which W is a $n \times n$ blockdiagonal matrix. Doing eigenvectors decomposition of the matrix K ($K = UDU^T$ where $U^T U = I$), simplifies λ to

$$\lambda = \frac{\beta^T U^T W U \beta}{\beta^T U^T U \beta} \tag{A.6}$$

where $\beta = DU^T\alpha$. Since U is orthonormal, Eq. A.6 can be simplified to $\lambda\beta = U^TWU\beta$. Where β corresponds to the d classical eigenvectors of the matrix U^TWU which after finding them, α_i can be computed by $\alpha = UD^{-1}\beta$ then normalized by

$$\alpha_i = \frac{\alpha_i}{\sqrt{\alpha_i^T K \alpha_i}} \tag{A.7}$$

at the end, projections of data points onto the normalized eigenvectors in the mapped space is found using α_i and function k like in KPCA.

A.2.3 Metric Multidimentional Scaling (MDS)

Metric Multidimensional Scaling (MDS) is another classical technique for dimensionality reduction. MDS tries to perform dimension reduction while preserving pairwise distances between data points as much as conceivable [52]. Technically, metric MDS tries to form new representation of data Y in space with lower dimensions by minimizing pairwise distances between data points in original and new space in compare to each other as

$$D = \sum_{i=1}^{N}\sum_{j=1}^{N}(d_{ij}^{(X)} - d_{ij}^{(Y)})^2 \tag{A.8}$$

where $d_{ij}^{(X)}$ and $d_{ij}^{(Y)}$ are the pairwise Euclidean distance between the data points in original space and in low dimensional space, respectively. The solution of this equation has been derived by eigendecomposition of the Gram matrix ($K = X^TX$) and it is equal to $Y = \Lambda^{1/2}V^T$ where V is the eigenvectors ofX^TX corresponding to the d leading eigenvalues, and Λ is the d leading eigenvalues of X^TX. In general, the distances between the pairs of data samples don't necessarily need to be Euclidean distances. It can also be many types of dissimilarities between samples such as preserving ordinal relations in data.

A.2.4 Stocastic Proximity Embedding (SPE)

Stochastic Proximity Embedding (SPE) proposed as an algorithm for iteratively generation of lower dimensions by minimizing the MDS raw stress function based on proximity data [53]. The transformation of data in lower dimensions with SPE is performed retaining the similarities between a set of related data points. SPE differs from MDS mainly because of the method it uses to update the low dimensional representation of data. SPE can also be utilized to preserve the distances in the neighborhood graph like the one comparable to ISOMAP. This algorithm iteratively minimizes the MDS stress function of

$$S = \sum_{ij} (d_{ij} - r_{ij})^2 \tag{A.9}$$

where r_{ij} and d_{ij} are the proximity between the high dimensional data points and the Euclidean distance between their low dimensional images, respectively [53]. Minimization is done in an iterative way and in four different steps:

1. The initialization of the points in the low dimensional subspace is performed randomly. Learning rate parameter λ is also selected.

2. An update procedure of the data points is carried out based on the random selection of s pairs of points in the low dimensional subspace and then computing their Euclidean distances and deriving the updated points by

$$x_i = x_i + \lambda \frac{1}{2} \frac{r_{ij} - d_{ij}}{d_{ij} + \varepsilon} (x_i - x_j) \tag{A.10}$$

$$x_j = x_j + \lambda \frac{1}{2} \frac{r_{ij} - d_{ij}}{d_{ij} + \varepsilon} (x_j - x_i) \tag{A.11}$$

   where ε is a small value preventing division by zero.

3. Learning rate λ decreases with the number of iterations by a predetermined level.

4. Repeating steps $2-3$ for a number of iterations, which is to update the embedded coordinates.

SPE has several advantages such as capability of dealing with dataset with large number of data points, no need to have the complete proximity matrix (r_{ij}), and quadratic scaling in memory and time which leads to the possibility of doing high number of iterations with a low computational costs.

A.2.5 Stocastic Neighbor Embedding (SNE)

Stochastic Neighbor Embedding (SNE) is a probabilistic iterative algorithm. This algorithm is used to map a high dimensional data representation onto a low dimensional representation based on preserving neighbor identities [55]. It means that in SNE, cost function is highly dependent on the similarities of nearby sample points. This is derived by centering a Gaussian on each high dimensional data point and calculating probability p_{ij} which represents the probability of data points x_i and x_j belonging to the same Gaussian as

$$p_{ij} = \frac{exp(-d_{ij}^2))}{\sum_{k \neq i} exp(-d_{ik}^2)} \tag{A.12}$$

where $d_{ij}^2 = \frac{\|x_i - x_j\|^2}{2\sigma_i^2}$ and σ^2 is a user-defined parameter that the locality behavior of SNE strongly depends on it. This calculation should be performed for all pairwise data points which form matrix P. Then, the low dimensional representation of data should be initialized randomly to the values very close to the origin and dimensionality of low dimensional subspace needs to be chosen much less than the number of data samples. The same as p_{ij}, probabilities q_{ij} for images after computation are stored in matrix Q. Difference between two probability distributions is measured using Kullback-Leibler divergences. Therefore, it can be used to minimize the difference between the two probability distributions of P and Q. In other word, SNE tries to minimize the natural

cost function of sum of Kulback-Leibler divergences given by

$$CF = \sum_i \sum_j p_{ij} log \frac{p_{ij}}{q_{ij}} = \sum_i KL(P_i \,||\, Q_i). \tag{A.13}$$

Minimizing this cost function using gradient decent method, local distances are emphasized so as similar data points placed close and dissimilar data points stay far apart in the low dimensional representation.

A.2.6 Local Linear Embedding (LLE)

Local Linear Embedding (LLE) is a nonlinear dimensionality reduction algorithm which works based on retaining the local properties of the data manifold. This has been carried out by an assumption that each data sample and its k-nearest points locate on or lie adjacent to a locally linear patch of data manifold [56]. After searching for k-nearest neighbors of each data sample, linear coefficients W_{ij} which reconstruct each data point from its k-nearest points (fitting a hyper plane through the data points) is found by using reconstruction errors in Eq. A.14 subject to $\sum_j W_{ij} = 1$ and $W_{ij} = 0$ if x_i and x_j are not in the same set of neighbors

$$arg \min_W E_W = \sum_i \left\| x_i - \sum_j w_{ij} x_{ij} \right\|^2, \tag{A.14}$$

where x_{ij} is the j_{th} neighbor of the i_{th} data point. These constrained coefficients of the data point x_i are invariant to rescaling, translation, and rotation. Due to these invariabilities, to preserve the local geometry after the linear mapping, these coefficients W_i can be used to reconstruct data point y_i in the lower dimensional subspace with the help of its neighbors. This is carried out via the embedding cost function of

$$arg \min_Y \; E_Y = \sum_i \left\| y_i - \sum_j w_{ij} y_j \right\|^2 \tag{A.15}$$

according to the unknown points y_i. Solving Eq. A.15, the coordinates of the representation of low-dimensional data y_i can be calculated solving an eigendecomposition of matrix $(I-W)^T(I-W)$.

A.2.7 Local Tangent Space Alignment (LTSA)

Local Tangent Space Alignment (LTSA) [88] is a nonlinear algorithm for dimensionality reduction based on describing the local data properties. In LTSA, due to presenting the local geometry of the manifold, local tangent space in the vicinity of each sample point is constructed. The local tangent spaces afterward are aligned to form the internal global coordinates for the nonlinear data manifold which LTSA assumes the local linearity of it. Based on this, a linear mapping of a data point in original space to its local tangent space and another one from its low-d image to the same local tangent space exist. Therefore, this algorithm tries to find this representation of data in low dimensions and the linear mappings of the low-d data samples onto the local tangent space. LTSA uses PCA on the k number of data samples (x_{ij}) which are neighbors of x_i. This is performed in order to make bases of the local tangent spaces at the same point x_i. This leads to mapping onto the local tangent space Θ_i from the neighborhood of x_i. Indeed existence of a linear mapping (L_i) is a property of Θ_i . This mapping is from the local tangent space coordinates (θ_{ij}) onto the low-dimensional representations (y_{ij}). Based on this, LTSA uses

$$\min_{Y_i,L_i} \|Y_i J_k - L_i \Theta_i\|^2 \tag{A.16}$$

which J_k would be the k sized centering matrix [47]. It is shown that using iterative summation for all V_i matrices would form the entries of an alignment matrix B, which begin from $b_{ij}=0$ for $\forall_{ij}$ as

$$B_{N_i N_i} = B_{N_i N_i} + J_k(I - V_i V_i^T)J_k \tag{A.17}$$

where N_i is a selection matrix. This matrix contains the indices of the data samples x_{ij}. The eigenvectors of an alignment matrix B relating to the least d nonzero eigenvalues forms the solution of Eq. A.16. Moreover, using the symmetric matrix $\frac{1}{2}(B + B^T)$ helps to find the low dimensional representation of data (Y).

A.2.8 ISOMAP

MDS is successful in many applications in pattern recognition, however it has a problem dealing with high-dimensional nonlinear manifolds. This problem emerges because it works using the Euclidean distance and doesn't take the distribution of neighboring data points in to account. ISOMAP as a nonlinear variant of the MDS algorithm is a nonlinear algorithm which tries doing dimensionality reduction based on the structure of manifold [54]. In fact, ISOMAP searches for an optimal subspace which highly retains the geodesic distances between data samples along the nonlinear manifold. The geodesic (or curvilinear) distances between two points along the manifold are the shortest paths. The algorithm aims to keep the pairwise geodesic distance between the sample points in original space close to the Euclidean distance between their low-dimensional images. According to the J.B Tenenbaum et al [54], the complete ISOMAP algorithm can be performed in three steps,

1. Constructing neighborhood graph finding neighbours of each data point
2. Computing the geodesic pairwise distances between all points
3. Constructing d-dimensional embedding via MDS

In this algorithm, every high dimensional data points should connect to its k nearest neighbors on the manifold constructing a neighborhood graph G. Then to form pairwise geodesic distance matrix D_G, the shortest path distance between all pairs of points in G can be calculate using Dijkstras algorithm [89] and Floyds algorithm [90]. At last, MDS is applied on D_G matrix to derive the low-dimensional coordinates y_is of Y that retains the estimated intrinsic geometry of the manifold in a d-dimensional space.

A.2.9 Laplacian Eigenmaps (LE)

Laplacian Eigenmaps (LE) is another algorithm which tries to conduct dimensionality reduction so that preserving local neighborhood information in the manifold [91]. This algorithm assumes that the data points lie on a smooth hyper plane. Then it computes the low dimensional representation of data using the intrinsic geometric structure of the manifold. Actually, LE tries to minimize the distance between each data and its nearest neighbor points. This is performed using a cost function containing the weighted distances between the higher dimensional data points so that the closer the points the bigger the weights. Like ISOMP in LE algorithm, every high dimensional data points are connected to its k nearest neighbors constructing a neighborhood graph G. For each edge , e_{ij}, that connects two points of the graph, a weight $W(i,j)$ is allocated which is a closeness-measure of the neighbor points as

$$W(i,j) = \begin{cases} exp(-\frac{\|x_i - x_j\|^2}{\sigma^2}) & if\ two\ points\ are\ neigbors \\ 0 & otherwise \end{cases} \tag{A.18}$$

where σ^2 is a user defined parameter. Calculation of all the weights forms matrix W. Minimizing the cost function by satisfying the closeness in low dimensional representation leads to

$$E = arg\,min \sum_{i=1}^{N} \sum_{j=1}^{N} (y_i - y_j)^2 W(i,j) \tag{A.19}$$

Based on this function, small distance between two points in high dimensional representation of data leads to the large weights. Thus to minimize the cost function, the corresponding distance between their low dimensional images should stay small too. Finally, the low dimensional representation of data (Y) are generated solving eigenequation $Lv = \lambda Dv$, where $L = D - W$ and D is a diagonal matrix with elements $D_{ii} = \sum_j W(i,j)$.

A.2.10 Hessian LLE

The Hessian LLE (HLLE) [92] is another nonlinear technique for dimensionality reduction. HLLE would be carried out while minimizing the manifolds curviness constraining the low dimensional representation of data being locally isometric. The manifold curviness is calculated by the local Hessian at each sample point (x_i). Indeed, the local Hessian in the local tangent space at the data point is tried to be represented. Then dimensionality reduction with HLLE is performed based on the eigenanalysis of Hessian matrix H [47]. In general, HLLE algorithm bears substantial resemblance to the LLE procedure and, similar it, at first identifies k-nearest neighbors (x_{ij}) of each data point x_i assuming the manifold is locally linear. By applying PCA on the neighborhood of k-nearest neighbors of x_i, and computing the d principal components from the covariance matrix $Cov\ x_{ij}$ and constructing U with them, a basis for the local tangent space is achieved. Then in order to compute an Hessian estimator of the manifold at the x_i, a matrix Z_i with elements equal to cross product of U up to the d_{th} order is formed. Subsequently, this matrix Z_i is orthonormalized with the use of Gram-Schmidt orthonormalization technique. After that, using the transpose of the last $d(d+1)/2$ columns of Z_i indeed can estimate the tangent Hessian (H_i). Then using Z_i, a symmetric matrix H can be calculated with elements,

$$H_{lm} = \sum_i \sum_j ((H_i)_{jl} \times (H_i)_{jm}) \tag{A.20}$$

The curviness of the manifold has been shown by this matrix H. At the end, Eigen-analysis of the H would result to the eigenvectors corresponding to the least d nonzero eigenvalues which form the new representation of data (matrix Y) so that curviness of the manifold is minimized.

A.3 Linear Techniques (Modified Nonlinear Techniques)

A.3.1 Locality Preserving Projections (LPP)

Locality Preserving Projections (LPP) is an algorithm for dimensionality reduction using a linear projective map. LPP works based on solving a variational problem like Laplacian Eigenmaps (LE) [56] which retains the local neighborhood information in the manifold [57]. LPP in fact is a linear approximation to LE and attempts to form a transformation matrix for projecting the data points into a space with lower dimensions. Unlike LE which simply works on training data and it is not clear in dealing with the images of test data points, LPP can be easily applied to new test data in order to map the points onto the low dimensional space. Moreover, LPP minimizes the cost function of a nonlinear algorithm (LE) which is different than the procedure in traditional linear techniques. So it is capable to take the nonlinearity of the manifold in to consideration by its locality preserving properties. Although LPP is a linear technique, the nonlinear structure of the manifold can be unraveled by it so that bypassing nonlinear expensive computations. Due to these advantages, when a linear technique is required for an accurate out of sample extension or de-embedding the data into original space, LPP can be utilized which is linear and solves the cost function of a local non-linear technique(LE). LPP aims at finding a transformation matrix A which helps to project the high dimensional points x_i to low dimensional y_i, as $y_i = A^T x_i$. Similar to LE, LPP constructs a neighborhood graph and matrix W, however the equation it uses is

$$arg \min_{a^T X D X^T a = 1} a^T X L X^T a \tag{A.21}$$

where L and D are the same as those in LE algorithm and x_i is the i_{th} column of matrix X. The vector a which minimizes Eq. A.21, is found by solving the eigenequation of

$$X L X^T a = \lambda X D X^T a. \tag{A.22}$$

In result, the embedded data would be $y_i = A^T x_i$, which $A = (a_0, ..., a_{d-1})$, and it is a $n \times d$ mapping matrix with column vectors of $a_0, ..., a_{d-1}$. These vectors are the eigenvectors of Eq. A.22 corresponding to its d least eigenvalues.

A.3.2 Neighborhood Preserving Embedding (NPE)

Neighborhood Preserving Embedding (NPE) [58] is another technique for dimensionality reduction which in fact is a linear approximation to the LLE algorithm [56]. Like LLE, NPE tries to preserve the local structure of manifold while doing data embedding. The most important advantage of NPE over LEE is its capability to naturally map the new testing samples as fast as training points using the transformation matrix A. In addition, comparing to classical linear techniques like PCA and FDA which work based on global structure, NPE is linear as well and it has neighborhood preserving properties. This feature allows it to discover nonlinear manifold bypassing nonlinear expensive computations. The algorithm starts like LLE, defining a neighborhood graph on high dimensional space and thereafter computing the reconstruction weights W_i. Each sample point x_i by a linear combination of the neighboring data points would be approximated. Then, an embedding which preserves the neighborhood structure would be performed in a lower dimensional space. This is done by optimizing the cost function of LLE and solving the following generalized eigenequation $XMX^T a = \lambda XX^T a$, where $M = (I - W)^T (I - W)$, $X = (x_1, \ldots, x_N)$, and "I" is the identity matrix. The column vectors $a_0, \ldots, a_{d-1}$ are the eigenvectors corresponding to the d least eigenvalues derived by this equation and form mapping matrix A. Thus, low dimensional data representation can be achieved using this $n \times d$ mapping matrix A.

A.3.3 Linear Local Tangent Space Alignment (LLTSA)

Linear Local Tangent Space Alignment (LLTSA) [93], similar to the linearization techniques in NPE and LPP, is a linear approximation of the nonlinear LTSA algorithm. This algorithm tries to preserve the local geometric structure in the vicinity of the data points with respect to the tangent

space. In fact, this algorithm defines a neighborhood graph on the data and the tangent space would be estimated in the vicinity of a data point x_i due to represent the local geometry of the manifold. Afterward the local tangent spaces are aligned and the alignment matrix , B, is formed. At the end, LLTSA attempts to form a matrix (A) for mapping X as dataset in high dimensions to their images in low dimensional space Y, such that $Y = (X - \bar{X})^T A$. The mapping matrix, A, is constructed by minimizing the LTSA cost function with a linear solution of the generalized eigenequation of $(X - \bar{X})^T B(X - \bar{X})a = \lambda(X - \bar{X})^T(X - \bar{X})a$. The d eigenvectors corresponding to the d least, nonzero eigenvalues forms the columns of this linear mapping matrix.

Appendix B

Classification Success Rates

Tables B.1 - B.12 present the classification success rate for various pairs of feature extraction/classification algorithms for each individual source of PD. The last column of each table shows the overall success rate.

Table B.1: Classification rate of classifiers on data output of **Statistical Operators**

	Class 1%	**Class 2**%	**Class 3**%	**Class 4**%	**Class 5**%	**Class 6**%	**Overall**%
SVM	100	100	100	100	90.0	62.5	92.1
KSVM	100	100	100	100	91.7	75.0	94.5
FSVM	100	100	100	100	93.3	80.0	95.6
FkNN	100	100	88.3	100	63.3	100	91.9
MLP	100	100	88.3	86.7	73.3	100	91.4
RBFN	100	98.3	90.0	98.3	85.0	100	95.3
PNT	100	100	93.3	96.7	71.7	100	93.6
Bayesian	100	100	83.3	98.3	88.3	98.3	94.7
Naive-B	100	100	100	90.0	50.0	91.7	88.6
AdaBoost	100	100	95.0	96.6	83.3	100	95.8

Class 1. Floating electrode in SF_6; **Class 2**. Point-plane electrodes in SF_6; **Class 3**. Free aluminum particle in SF_6; **Class 4**. Free aluminum particle in oil; **Class 5**. Point-plane electrodes in oil; **Class 6**. Point-plane electrodes in air.

Table B.2: Classification rate of classifiers on data output of **PCA**

	Class 1%	**Class 2%**	**Class 3%**	**Class 4%**	**Class 5%**	**Class 6%**	**Overall%**
SVM	100	98.3	88.3	100	87.5	100	95.7
KSVM	100	100	95.0	100	90.0	100	97.5
FSVM	100	100	96.7	100	90.0	100	97.8
FkNN	100	100	96.7	100	65.0	100	93.6
MLP	100	78.3	100	100	81.6	100	93.3
RBFN	100	100	100	98.3	86.7	96.7	96.9
PNT	100	88.3	100	100	83.3	100	95.3
Bayesian	100	100	98.3	98.3	80.0	100	96.1
Naive-B	100	50.0	98.3	100	86.7	100	89.2
AdaBoost	100	100	96.6	100	88.3	100	97.5

Table B.3: Classification rate of classifiers on data output of **FDA**

	Class 1%	**Class 2%**	**Class 3%**	**Class 4%**	**Class 5%**	**Class 6%**	**Overall%**
SVM	100	100	92.5	100	84.2	100	96.1
KSVM	100	100	96.7	100	91.7	100	98.1
FSVM	100	100	96.7	100	91.7	100	98.1
FkNN	100	91.7	98.3	100	88.3	96.7	95.8
MLP	100	95.0	90.0	100	98.3	100	97.5
RBFN	100	100	96.7	100	90.0	100	97.7
PNT	100	100	91.7	100	96.7	100	98.1
Bayesian	100	100	96.7	100	86.7	100	97.2
Naive-B	100	98.3	96.7	100	75.0	100	95.0
AdaBoost	100	100	96.7	100	95.0	100	98.6

Table B.4: Classification rate of classifiers on data output of **Kernel FDA**

	Class 1%	**Class 2%**	**Class 3%**	**Class 4%**	**Class 5%**	**Class 6%**	**Overall%**
SVM	96.6	100	98.3	100	95.0	61.6	91.9
KSVM	95.0	100	96.6	100	96.6	83.3	95.2
FSVM	95.0	100	98.3	100	96.6	90.0	96.7
FkNN	93.3	100	93.3	100	90.0	100	96.1
MLP	98.3	100	98.3	100	75.0	78.3	91.7
RBFN	96.6	100	100	100	80.0	98.3	95.8
PNT	98.3	100	98.3	100	76.6	83.3	92.8
Bayesian	90.0	100	93.3	100	95.0	100	96.4
Naive-B	96.7	100	93.3	98.3	83.3	78.3	91.7
AdaBoost	98.3	100	91.7	100	98.3	98.3	97.8

Table B.5: Classification rate of classifiers on data output of **KPCA**

	Class 1%	**Class 2%**	**Class 3%**	**Class 4%**	**Class 5%**	**Class 6%**	**Overall%**
SVM	100	98.3	81.7	100	88.3	100	94.7
KSVM	100	100	93.3	100	95.0	100	98.1
FSVM	100	100	95.0	100	98.1	100	98.9
FkNN	100	100	100	98.3	87.5	100	97.6
MLP	98.3	95.0	95.0	100	98.3	100	97.7
RBFN	98.3	100	98.3	100	95.0	100	98.6
PNT	100	100	76.6	100	100	100	96.1
Bayesian	100	100	98.3	100	80.0	100	96.4
Naive-B	100	98.3	53.3	100	90.0	100	90.3
AdaBoost	100	100	93.3	100	93.3	100	97.8

Table B.6: Classification rate of classifiers on data output of **MDS**

	Class 1%	**Class 2%**	**Class 3%**	**Class 4%**	**Class 5%**	**Class 6%**	**Overall%**
SVM	100	100	88.3	100	85.0	100	95.5
KSVM	100	100	95.0	100	100	100	99.1
FSVM	100	100	96.7	100	100	100	99.4
FkNN	100	93.3	100	100	98.3	100	98.6
MLP	98.3	100	96.7	100	91.6	100	97.7
RBFN	100	58.3	83.3	100	100	100	96.9
PNT	100	100	78.3	100	100	100	96.4
Bayesian	100	98.3	90.0	100	98.3	100	97.8
Naive-B	100	100	86.7	100	81.7	96.7	94.2
AdaBoost	100	96.6	95.0	100	95.0	98.3	97.5

Table B.7: Classification rate of classifiers on data output of **SPE**

	Class 1%	**Class 2%**	**Class 3%**	**Class 4%**	**Class 5%**	**Class 6%**	**Overall%**
SVM	83.3	91.6	80.0	100	50.0	100	84.2
KSVM	90.0	85.0	85.0	100	73.3	100	88.9
FSVM	91.6	88.3	85.0	100	78.3	100	90.5
FkNN	98.3	83.3	40.0	100	56.6	100	79.7
MLP	91.6	93.3	71.6	100	75.0	100	88.6
RBFN	88.3	90.0	73.3	100	85.0	100	89.3
PNT	100	91.6	68.3	100	60.0	90.0	85.0
Bayesian	100	83.3	83.3	98.3	78.3	98.3	90.2
Naive-B	65.0	70.0	90.0	100	40.0	91.7	76.1
AdaBoost	91.7	78.3	88.3	98.3	83.3	98.3	89.7

Table B.8: Classification rate of classifiers on data output of **Isomap**

	Class 1%	**Class 2%**	**Class 3%**	**Class 4%**	**Class 5%**	**Class 6%**	**Overall%**
SVM	100	98.3	90.0	100	96.7	98.3	97.2
KSVM	98.3	100	96.7	100	96.7	100	98.6
FSVM	98.3	100	96.7	100	96.7	100	98.6
FkNN	100	100	100	100	78.3	100	96.4
MLP	100	83.3	98.3	100	100	100	96.9
RBFN	100	100	98.3	98.3	85.0	75.0	92.8
PNT	100	100	100	100	91.6	85.0	96.1
Bayesian	96.7	91.7	96.7	96.7	100	100	96.9
Naive-B	100	68.3	100	98.3	68.3	100	89.2
AdaBoost	100	100	95.0	100	91.7	98.3	97.5

Table B.9: Classification rate of classifiers on data output of **SNE**

	Class 1%	**Class 2%**	**Class 3%**	**Class 4%**	**Class 5%**	**Class 6%**	**Overall%**
SVM	98.3	98.3	93.3	100	93.3	91.7	95.8
KSVM	98.3	100	95.0	100	96.7	95.0	97.5
FSVM	100	100	95.0	100	96.7	98.3	98.3
FkNN	100	95.0	100	100	98.3	100	98.8
MLP	100	100	100	100	96.6	75.0	95.2
RBFN	96.7	98.3	95.0	100	90.0	78.3	93.1
PNT	100	100	100	100	93.3	68.3	93.6
Bayesian	98.3	90.0	90.0	100	91.7	96.7	94.4
Naive-B	98.3	93.3	78.3	100	88.3	95.0	92.2
AdaBoost	100	98.3	95.0	100	95.0	98.3	97.8

Table B.10: Classification rate of classifiers on data output of **LLE**

	Class 1%	**Class 2%**	**Class 3%**	**Class 4%**	**Class 5%**	**Class 6%**	**Overall%**
SVM	98.3	100	91.7	100	83.3	100	95.5
KSVM	100	100	93.3	100	91.6	100	97.5
FSVM	100	100	93.3	100	93.3	100	97.8
FkNN	98.3	93.3	96.7	68.3	75.0	70.0	83.6
MLP	100	50.0	41.6	100	96.7	100	81.4
RBFN	100	100	73.3	100	96.7	100	95.0
PNT	100	50.0	48.3	100	98.3	100	82.7
Bayesian	100	98.3	98.3	100	80.0	100	96.1
Naive-B	100	100	73.3	100	95.0	100	94.7
AdaBoost	100	91.7	100	100	96.7	100	98.1

Table B.11: Classification rate of classifiers on data output of **NPE**

	Class 1%	**Class 2%**	**Class 3%**	**Class 4%**	**Class 5%**	**Class 6%**	**Overall%**
SVM	100	65.0	98.3	70.0	96.7	100	88.3
KSVM	98.3	98.3	98.3	96.7	91.7	93.3	96.1
FSVM	98.3	98.3	100	98.3	91.7	93.3	96.7
FkNN	93.3	100	98.3	90.0	95.0	96.3	95.5
MLP	98.3	96.6	90.0	73.3	90.0	95.0	90.5
RBFN	90.0	96.6	81.7	88.3	93.3	95.0	90.8
PNT	83.3	100	58.3	100	88.3	98.3	88.1
Bayesian	100	90.0	95.0	85.0	90.0	86.7	91.1
Naive-B	93.3	90.0	86.7	51.6	60.0	90.0	78.6
AdaBoost	100	96.7	93.3	90.0	95.0	91.7	94.4

Table B.12: Classification rate of classifiers on data output of **LPP**

	Class 1%	**Class 2%**	**Class 3%**	**Class 4%**	**Class 5%**	**Class 6%**	**Overall%**
SVM	100	85.5	86.7	100	88.3	100	93.3
KSVM	100	96.7	91.7	100	93.3	100	96.9
FSVM	100	98.3	93.3	100	96.7	100	98.1
FkNN	100	63.3	98.3	100	96.7	100	93.1
MLP	100	85.0	100	100	75.0	98.3	93.1
RBFN	100	98.3	96.6	100	91.7	100	97.8
PNT	100	80.0	98.3	100	98.3	100	96.1
Bayesian	100	85.0	93.3	98.3	98.3	100	95.8
Naive-B	100	53.3	100	100	100	98.3	91.9
AdaBoost	100	100	90.0	98.3	95.0	98.3	96.9