

Towards automated statistical partial discharge source classification using pattern recognition techniques

ISSN 2397-7264
Received on 1st June 2018
Revised 17th July 2018
Accepted on 18th July 2018
E-First on 6th September 2018
doi: 10.1049/hve.2018.5048
www.ietdl.org

Hamed Janani^{1,2}, Behzad Kordi¹ ✉

¹Department of Electrical and Computer Engineering, University of Manitoba, 75A Chancellor's Circle, Winnipeg, MB, Canada, R3T5V6

²PowerTech Labs Inc., 12388-88 Avenue, Surrey, BC, Canada, V3W 7R7

✉ E-mail: Behzad.Kordi@umanitoba.ca

Abstract: This study presents a comprehensive review of the automated classification in partial discharge (PD) source identification and probabilistic interpretation of the classification results based on the relationship between the variation of the phase-resolved PD (PRPD) patterns and the source of the PD. The proposed automated classification system consists of modern, high-performance statistical feature extraction methods and classifier algorithms. Their application in online monitoring and recognition of the PD patterns is investigated based on their low-processing time and high-performance evaluation. The application of modern statistical algorithms and pre-processing methods configured in this automated classification system improves the pattern recognition accuracy of the different PD sources that are suitable to be employed in different high-voltage (HV) insulation media. To evaluate the performance of the different combinations of the feature extraction/classifier pairs, laboratory setups are designed and built that simulate various types of PDs. The test cells include three sources of PD in SF₆, two sources of PD in transformer oil, and corona in the air. Data samples for different classes of PD sources are captured under two levels of voltage and two different levels of noise. The results of this study evaluate the suitability of the proposed classification systems for probabilistic source identification in various insulation media. Furthermore, of importance to the problem of the PD source identification is to assign a 'degree of membership' to each PRPD pattern, besides assigning a class label to it. Some of the classifier algorithms studied in this study, such as fuzzy classifiers, are not only able 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 determination of the degree of membership for future PRPD samples allows safer decision making based on the risk associated with the different sources of PD in HV apparatus.

1 Introduction

Safe, stable, and reliable electric power systems rely on solid, liquid, and gaseous insulation materials to isolate energised 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, resulting in a gradual 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 (PDs) 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 PDs, as a symptom of insulation deterioration, can be used to improve the reliability of HV insulation. Early detection of PDs and their cause prevents costly failures of electrical equipment. PD measurements were first carried out almost 80 years ago [2, 3], but were not considered seriously for the reliability assessment of HV insulations until the 1950s–1960s [4]. The techniques employed for PD detection are based on chemical, acoustic, optical, electrical, and ultra-high frequency measurements. The electrical measurements of the PD in HV alternating current (AC) systems are widely used and are the focus of this work.

One important application of the PD measurements is the identification of the source of the PD. In the AC systems, the phase-resolved PD (PRPD) pattern, which visualises the occurrence of the PD activities in reference to the phase of the 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. This forms a bivariate distribution where each of the discharge magnitude and discharge rate can be separately analysed 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 the PRPD patterns by human experts has been one of the major PD analysis approaches. The identification of defects is normally accompanied by uncertainty due to some overlap of different discharge patterns. However, in recent years, due to the availability of high-speed data processors and well-developed statistical techniques in machine learning, the automated identification of the PD sources seems to be more achievable [6]. Reliable, automated, and classification of the PD sources enables online monitoring of 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).

The existence of a correlation between the nature of the PD sources and their PRPD patterns has been the motivation of designing a thoroughly automated feature extractor and pattern classifier system for the application in the area of HV insulation monitoring. In the last three decades, the automated recognition of the PD patterns has been progressively investigated and several signal processing methods and classification algorithms have been employed for the analysis of discharge patterns, such as the relative identification factor [7], time series analysis [8, 9], artificial neural networks (ANNs) [4, 5, 9–13], fuzzy algorithms [14], support vector machine (SVM) [15, 16], hidden Markov models [17], pattern recognition based on the chaos theory [18], kernel statistical uncorrelated optimum discriminant vectors algorithm

[19], statistical tools [6, 20], inductive learning approach [21], Bayesian [15], and K means [22]. In [6], Krivda used Fisher discriminant analysis (FDA) and principal component analysis (PCA) for feature extraction. The application of the wavelet transforms has also been shown to be useful for PD source recognition [12, 15]. To improve the performance of the ANN in the classification of discharge patterns, the knowledge-based pre-processing method and time series approach have been presented in [8, 10]. A characteristic that distinguishes some of these approaches from the others is the ability to assign a posterior probability to an unknown sample. Such a characteristic enables risk management and decision-making for asset managers with regard to condition-based maintenance.

This study presents a review of the application of artificial intelligence and pattern recognition techniques to develop an automated classification system for the improvement of the PD source identification using modern statistical methods to identify PD sources with a high accuracy rate. The main characteristic sought is the capability of performing probabilistic interpretation of the classification results and calculating the ‘degree of membership’ of a sample to a class of data. This is a feature that simple techniques, such as k-nearest neighbour (kNN) [23], are not capable of. Another limitation of a kNN classifier is that it is a local approach that does not provide a model. kNNs tend to have low biases but high variances. They are also computationally expensive for large data sets and are unable to determine the importance of the features. Also, for high-dimensional cases, the notion of the nearest neighbours is very hard to define and the data tends to be sparse.

In this study, we explored techniques that are able to provide the statistical interpretation of the new data that are being classified. Knowing the posterior probability level of a sample is mostly important for the class prediction of a new, unknown PRPD sample which is generated from the same type of defect but does not originally belong to the original data set. This also enables us to take into account the risk associated with various types of defects by setting a threshold for an acceptable ‘degree of membership’ and referring marginal classifications to an expert operator.

An automated classification system consists of feature extraction methods and classifier algorithms that are implemented and suitable to improve the recognition of the PD patterns of various sources in different insulation media. Once the phase-resolved PD patterns are recorded, features are generated. Subsequently, to present a comprehensive classification system to work in different insulation media, high performance, applicable dimensionality reduction methods are chosen (exploring almost all available well-developed feature extraction techniques) that are combined with almost all well-known classifier algorithms. Dimensionality reduction is required to extract features that represent the fingerprints of the PD source. This removes redundant and ineffective information and decreases the number of features while still capturing a high portion of information [23, 24]. The feature extraction step is very important in this procedure and the efficiency of a classifier is highly dependent on the wellness of extracted features.

In this work, in order to present a comprehensive classification system and to explore almost all available well-developed feature extraction techniques, the application of 12 high performance dimensionality reduction techniques (including the traditional statistical operators) that are applicable to the PRPD pattern data is investigated. Following the feature extraction procedure, ten well-known algorithms for the classification of the PD sources are employed and investigated. The classification success rate of their application on the PD patterns of the discharge activities in different insulation media including air, oil, and SF₆ gas has been evaluated.

Some of the classifier algorithms studied such as fuzzy classifiers are not only able to show a high-classification accuracy rate but also can 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 feature for future PRPD samples would allow safer decision making based on the risk associated with the

different sources of the PD in HV apparatus. This also allows us to reject a sample from classification by setting a threshold for an acceptable ‘degree of membership’ for different sources in different HV insulation systems.

The paper is organised as follows. In Section 2, the framework that has been used for the classification of the PD sources is described. The experimental setup that was used to generate the datasets is presented in Section 3. The PRPD patterns of the test cells studied are presented in Section 4. Sections 5 and 6 are a review of the feature extraction techniques and classifiers. The results are discussed in Section 7, and finally, Section 8 concludes the paper.

2 Comprehensive framework for partial discharge classification

A pattern classification algorithm generally consists of three main steps [23, 24]:

- i. data pre-processing,
- ii. feature extraction (dimensionality reduction), and
- iii. implementation of the classifier algorithm and conducting a probabilistic interpretation.

Below is a description of each step when applied to a PD source classification problem using the PRPD patterns.

2.1 PRPD data pre-processing

PRPD patterns provide a bivariate distribution $H_n(\varphi, q)$ that shows the relationship 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. 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 derived. Considering these parameters in reference to the phase angle results in three univariate distributions of peak discharge, $H_{qmax}(\varphi)$, average discharge, $H_{qmean}(\varphi)$, and discharge rate, $H_n(\varphi)$, respectively. To generate one data point of the dataset, the PRPD pattern is recorded for T seconds and then the univariate distributions are evaluated. To generate a dataset of P points, we have to repeat this process P times. Finally, the data points are transformed into a matrix whose dimension is $3M \times P$. In this work, typical values for these parameters are $M = 100$, $P = 300$, and $T = 3$ s (or 180 cycles of a 60 Hz sinusoidal) for each type of defect used for training and evaluating the classifiers. The values of parameters M , P , and T are selected based on the optimisation algorithm that uses both the percentage of misclassified test samples as an estimate of the error rate and the processing time. This algorithm optimises the parameters by minimising the multiplication of these two factors. We normally allow 2 s between every two consequence data points.

2.2 PRPD feature extraction

The first major problem in building a classifier is the curse of dimensionality [23], which should be resolved by selecting a proper combination of available features through the application of dimensionality reduction. Another reason for reducing the features is due to 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. In addition, more features will make training a classifier more difficult. Moreover, the implementation of the feature extraction techniques leads to the removal of multi-collinearity, which improves the performance of the classification algorithm [23]. 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 the classification of the PDs in the past [6,15];

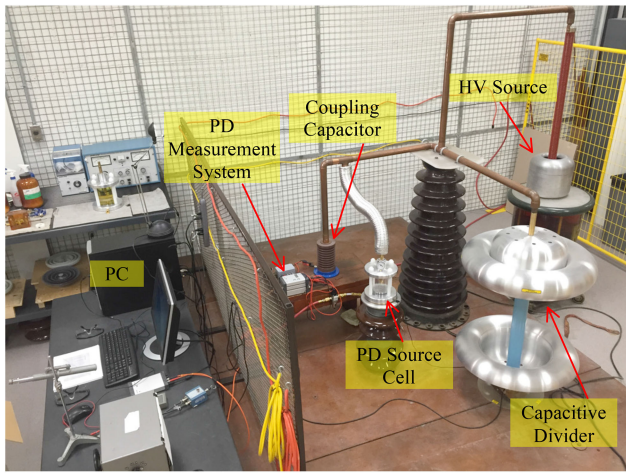


Fig. 1 Experimental test setup consisting of a 60-Hz HV source, a capacitive divider to measure the AC voltage, the test cell, and the Omicron MPD 600 PD measurement system

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 modern, more powerful techniques, which are suitable for improving the online monitoring of PD sources, are employed in the proposed system. These techniques attempt to extract and identify data resting on a low-dimensional manifold of dimension K ($K < 3M$), from a high-dimensional space \mathbb{R}^{3M} that the manifold is embedded in. ' K ' is typically referred to as the intrinsic dimension of the dataset [23].

2.3 PRPD 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 performing of two tasks: training (learning) and testing (classifying) [25].

The 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. Indeed, a recognition system is designed to assign future samples that are probably different than the training data. The trained system should be efficiently optimised to show the desired performance on the prediction of the test data. A highly-optimised 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 the classification of the 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 [25]. Moreover, the ratio of the number of training samples to the number of features is an important factor. If too small, it would influence the performance of the classifier (i.e. curse of dimensionality).

To design a powerful classifier for accurate PD source identification in different insulation media, a thorough investigation is required, using various algorithms for extracting features from the PRPD patterns and building a number of classifiers. This investigation is performed in this work to improve the classification accuracy rate of the PD sources in different insulation systems and conducting probabilistic interpretation of the results.

3 Experimental procedure

The classification of the different sources of the PD requires a database for training the classifier and testing. In almost all previous studies, such a database is generated based on the measurements conducted on artificial defects that are implemented in controlled laboratory test cells [4–6,11,14,15]. The measurement procedure and system calibration have been performed according

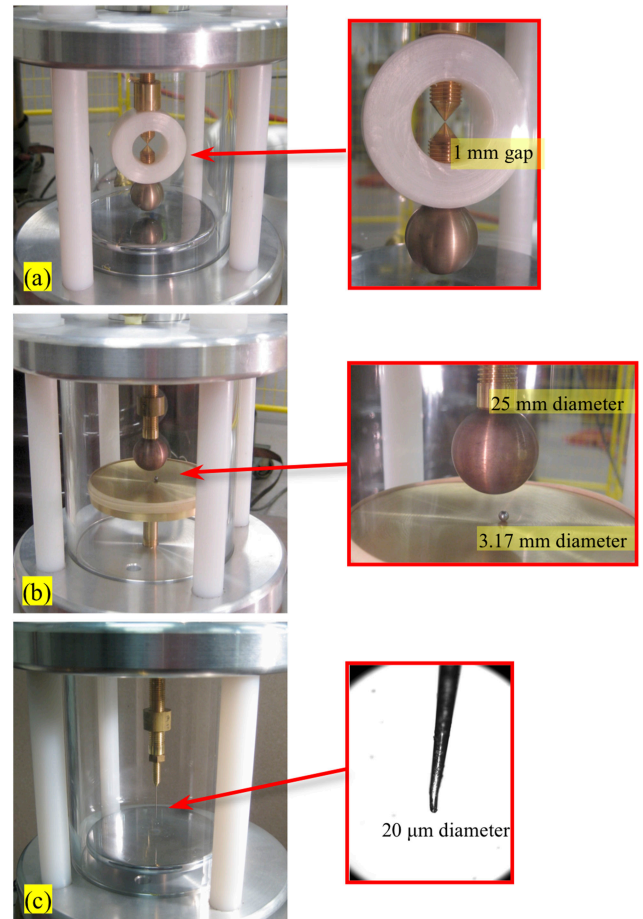


Fig. 2 SF₆ test cells

(a) Floating electrode, (b) Free particle, (c) Point-plane electrodes. Each cell consists of a Perspex tube clamped by nylon screws between the top and bottom aluminium caps that can withstand a pressure of up to 500 kPa

to the IEC 60270 standard [26] using a commercial PD measurement system, Omicron MPD 600.

Fig. 1 shows the experimental setup that consists of a HV transformer energising the test cell, a coupling capacitor, the quadruple measuring impedance, and commercial PD measuring equipment. The 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. The scope of the current work is limited to single-source defects. The authors have also proposed a novel approach to identify PD sources when multiple sources of PD are present [27].

The test cells used in this work, originally proposed by Hampton and Meats [28], are shown in Figs. 2 and 3. These test cells are built to model different types of PD activities with different discharge mechanisms in the air, oil, or SF₆. SF₆ test cells are designed to model the common defects of the GIS in small scale and be able to withstand a pressure of up to 500 kPa, consistent with gas pressure in the GIS. Sparking from a floating electrode, moving particles and the fixed protrusion are some of the major sources of the PD in a GIS [29] whose laboratory models are shown in Fig. 2. The test cells that can generate PD due to the free particles in oil and a needle electrode in oil are shown in Fig. 3. The same setup as that shown in Fig. 2c (but filled with air at 100 kPa) was employed to generate PD due to the corona in the air.

4 PRPD patterns of test cells

The PRPD patterns of the six PD test cells are shown in Fig. 4. For the floating electrode in SF₆ (see Figs. 2a and 4a), an inception voltage of ~15 kV was measured at 400 kPa. It was observed that the inception voltage and discharge magnitudes both increased with an increase in SF₆ pressure. It was also observed that both the inception voltage and PD magnitudes are strongly related to the

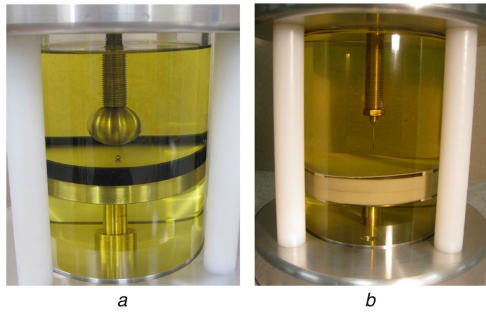


Fig. 3 Oil test cells
(a) Free particle, (b) Point-plane electrodes where the tip of the needle is 20 μm in diameter and the ground plane is covered with insulation paper to avoid flashover

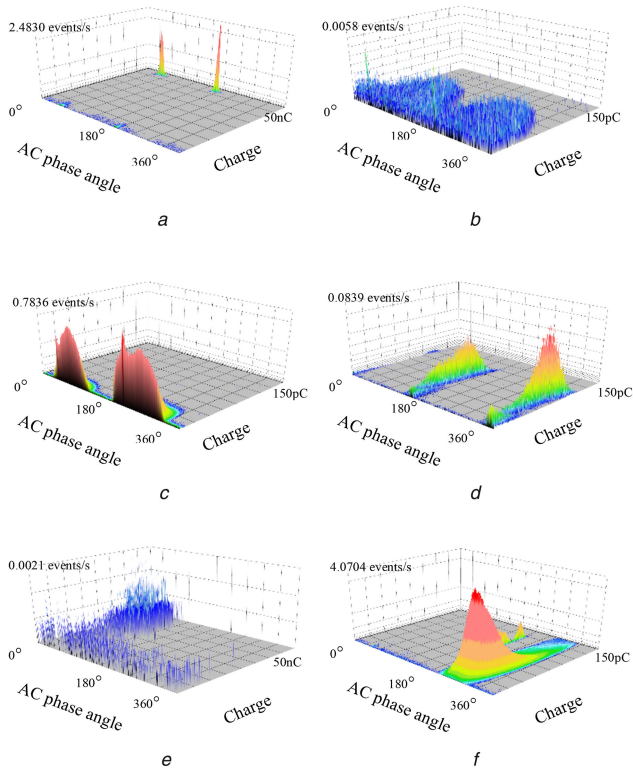


Fig. 4 PRPD patterns of
(a) Floating electrode in SF_6 , (b) Free particle in SF_6 , (c) Point-plane electrodes in SF_6 , (d) Free particle in oil, (e) Point-plane electrodes in oil, (f) Point-plane electrodes in air

gap size between the energised 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. 4b 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 10 mm from the ground electrode. As the voltage is increased to 10.5 kV, the small bearing starts to move across the plate towards the edge and back. This movement generates PDs between the bearing and ground dish. PDs occur because of the charges that are transferred from the bearing to the ground electrode [29]. This experiment was repeated for different sizes of the bearing. When the size of the 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 400 kPa is shown in Fig. 4c. To generate a PD in this setup a tungsten needle with a tip radius of 10 μm , located at a distance of 15 mm

from the ground plate, has been used. This corona pattern is, in fact, the PRPD pattern of 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 . The typical PD magnitude of the 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. Since the low level of discharge of negative corona, in this work, we have considered positive corona only. Positive corona inception voltage was measured at $\sim 15 \text{ kV}$ for this setup. However, the discharge magnitude remains almost the same while the applied voltage is in the range of 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 the PRPD pattern features with a variation in the SF_6 pressure.

The PRPD pattern of a free particle in oil is shown in Fig. 4d. This setup and the electrodes and distances are the same as the setup of Fig. 2b (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 the bearing and the ground plane. The PD leads to the release of gas bubbles which move from the PD location towards the HV electrode. Sometimes, 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. The larger bearing can store and transfer more charge to the ground electrode so PD magnitude becomes higher. Bouncing of the free particle under the HV electrode will also be more 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 the discharge region.

To model corona discharge in oil, another source of PD in oil, a 10 μm tungsten needle electrode configuration is used (Fig. 3b). The HV is connected to the needle with its tip located 10 mm away from the grounded electrode. To avoid flashover, a grounded electrode is covered with a piece of insulating paper. The PRPD pattern of the needle electrode is shown in Fig. 4e. The inception voltage of this test cell was 20 kV. It is observed that the PD effects in this pattern are more vigorous in the positive half cycle with a large dispersion.

The last PRPD pattern shown in Fig. 4f is related to the corona in the 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.

5 Feature extraction techniques for PRPD patterns

For datasets such as the PD dataset with a large number of features and a limited number of observations, feature extraction should be applied [23, 24]. Having more information about a PRPD pattern seems to be useful; however, having many features compared with the number of observations is not efficient for producing the 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. In this work, various computer codes are developed to employ modern feature extraction techniques which are suitable for online monitoring of PD sources.

5.1 Dimension reduction algorithms

A dimension reduction technique is a transformation method that transforms the data from the high-dimensional feature space to a new informative space with lower dimensionality [30]. In other words, such transformation transforms the matrix $\mathbf{X}_{3M \times NP}$ to a matrix $\mathbf{Y}_{K \times NP}$, 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 previous studies, a very limited number of dimension reduction techniques have been applied for PD recognition.

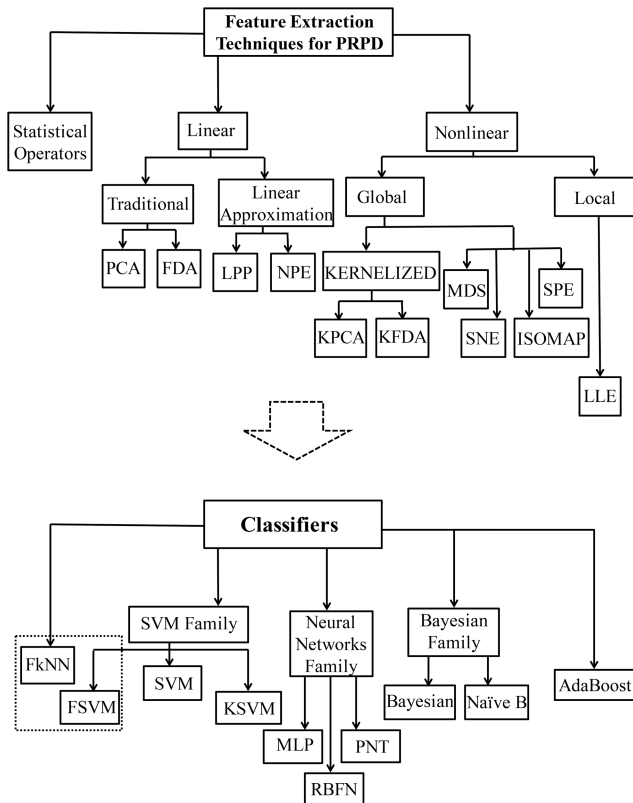


Fig. 5 PRPD feature extraction and classifier algorithms

However, improved PD source identification requires more modern and powerful techniques. In this work, all of the well-developed dimension reduction techniques were investigated, of which, 11 of them were found to be efficiently applicable to PRPD datasets. These techniques are divided into two main groups:

- i. linear techniques that include PCA [23], FDA [31], and
- ii. nonlinear techniques such as kernel PCA (KPCA) [32], kernel FDA (KFDA) [33], metric multidimensional scaling (MDS) [34], stochastic proximity embedding (SPE) [35], isomap [36], stochastic neighbour embedding (SNE) [37], and local linear embedding (LLE) [38].

A third group can also be identified under the linear group that is linear algorithms derived based on the linear approximation of some local nonlinear algorithms. These algorithms include linearity preserving projection (LPP) [39] and neighbourhood preserving embedding (NPE) [40]. A summary of the dimensionality reduction techniques is shown in Fig. 5.

The application of all these algorithms on the dataset that is generated from the PRPD patterns of the different sources of the PD has been performed and the results are fed to the next step of the machine learning algorithm after passing them through a pre-processing stage.

5.2 Statistical operators

Another approach to extract features which are capable of differentiating between the discharge patterns of different PD sources is to use statistical parameters which can be applied on the univariate distributions. Statistical operators that have been widely used in the literature (e.g. [5, 11]) for PRPD classification include mean, variance, skewness, Kurtosis, number of local peaks, discharge asymmetry, phase asymmetry, cross-correlation factor, and modified cross-correlation factor. 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 [23]. 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 a modified cross-correlation factor are in this group [23].

In this study, a novel approach with the discriminatory capability to separate different PD classes is implemented using the q -quantiles [23] in addition to the other statistical operators. In this approach, we divide the PD distribution into $q + 1$ groups with equal numbers of data points (in this work, $q = 3$ is assumed, i.e. PD distribution is divided into 0–25, 25–50, 50–75 and 75–100% of the total number of data points). By adding the extracted features using the q -quantiles to those obtained using the operators in the literature applied to $H_{qmax}(\varphi)$, $H_{qmean}(\varphi)$, and $H_n(\varphi)$, a fingerprint of each discharge pattern is generated. This approach considerably increases the discriminatory power of commonly-used statistical operators [6, 11, 15].

The results of the classifiers using statistical operators will be compared with classifiers that use dimension reduction techniques. A comparison of the overall classification success rate related to the specific feature extraction/classification algorithms will help in finding the more efficient combination of algorithms in different insulation media.

6 Pattern classification algorithms for PRPD patterns

Neural networks and SVM are the commonly-used classifier algorithms in PRPD recognition. However, we show that there are other techniques that have more superior performance, and in addition, are capable of estimating posterior probabilities. In this work, ten well-known algorithms for the classification of the PD sources have been used. These algorithms are: SVM [23], kernel SVM (KSVM) [23], fuzzy SVM (FSVM) [41], fuzzy kNN (FkNN) [42], multi-layer perceptron (MLP) [43], radial bases function networks (RBFN) [24, 43], probabilistic neural networks (PNN) [23, 44], Bayesian [23, 24], Naïve Bayes [23], and AdaBoost [45]. Some of these algorithms including fuzzy classifiers are not only capable of showing high classification accuracy rate, but they 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 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 the PD in HV apparatus. The fuzzy algorithms which have been used in this study are FkNN [42] and FSVM [41]. A summary of the classifier algorithms is shown in Fig. 5. Various computer codes are developed to employ the proposed feature extraction/classifier algorithms for online monitoring of PD sources. The performance evaluation of all classifier algorithms integrated with different feature extraction algorithms on PD source identification is presented in the following sections.

7 Results and discussions

7.1 Classification procedure

Using the experimental setups, a total of 300 data points are generated for each of the six different classes of the PD source. For each type of defect, the data points form the dataset matrix $X_{3M \times NP}$ whose dimension is 300×1800 (i.e. $M = 100$, $P = 300$, $N = 6$). The application of the dimension reduction algorithms listed in Section 5.1 (except for FDA and KFDA) on matrix X results in a dimension reduction from 300 in the original space to nine 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 [23]. However, for FDA and KFDA, the new dimension is equal to $K = 5$. This number is selected based on FDA and KFDA 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

Table 1 Overall classification accuracy rate of classifiers on data output of different feature extraction techniques

	Statistical operators	PCA	FDA	KFDA	Kernel PCA	MDS	SPE	Isomap	SNE	LLE	NPE	LPP
SVM	92.1	95.7	96.1	91.9	94.7	95.5	84.2	97.2	95.8	95.5	88.3	93.3
KSVM	94.5	97.5	98.1	95.2	98.1	99.1	88.9	98.6	97.5	97.5	96.1	96.9
FSVM	95.6	97.8	98.1	96.7	98.9	99.4	90.5	98.6	98.3	97.8	96.7	98.1
FkNN	91.9	93.6	95.8	96.1	97.6	98.6	79.7	96.4	98.8	83.6	95.5	93.1
MLP	91.4	93.3	97.5	91.7	97.7	97.7	88.6	96.9	95.2	81.4	90.5	93.1
RBFN	95.3	96.9	97.7	95.8	98.6	96.9	89.3	92.8	93.1	95.0	90.8	97.8
PNN	93.6	95.3	98.1	92.8	96.1	96.4	85.0	96.1	93.6	82.7	88.1	96.1
Bayesian	94.7	96.1	97.2	96.4	96.4	97.8	90.2	96.9	94.4	96.1	91.1	95.8
Naïve-B	88.6	89.2	95.0	91.7	90.3	94.2	76.1	89.2	92.2	94.7	78.6	91.9
AdaBoost	95.8	97.5	98.6	97.8	97.8	97.5	89.7	97.5	97.8	98.1	94.4	96.9

Bold values indicate the largest and second largest values.

KFDA), FDA (including KFDA), and statistical operators reduce the dimension of the original datasets from 300 to 9, 5, and 55, respectively. (Applying the five operators of mean, variance, skewness, Kurtosis, and number of peaks to both positive and negative cycles of $H_{qmax}(\varphi)$, $H_{qmean}(\varphi)$, and $H_n(\varphi)$ results in 30 features for each PRPD pattern. Furthermore, applying additional operators of discharge asymmetry, phase asymmetry, cross-correlation factor, and modified cross-correlation factors will generate extra seven features [6]. In addition, the application of the three 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.)

To perform the PD source classification, the dimension-reduced dataset Y is fed to the classifier algorithms for both training and testing purposes. To do a 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 optimise the different classifier parameters, first the data in 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 trade-off; if the training set becomes small, the classifier will not be very robust and if the test subset becomes small then the confidence in the estimated error rate will be low [25]. To run the optimisation procedure for different classifier parameters, a ten-fold cross validation (rotation method) is applied to the 80% training set. The n -fold cross-validation algorithm has been selected over leave-one-out or holdout methods because of its higher efficiency and better performance on the PD subset [25]. This method divides the training set into n subsets of equal size and 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 study, this cross-validation process has been repeated ten times and ten 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 optimisation, the classifier is trained using the whole training subset with the optimal value of parameters. In 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 optimisation 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 five times. Classification accuracy is averaged over the five trials and represents the success rate of each feature extraction/classifier algorithm. Data samples for the six different classes of PD sources are captured under two levels of voltage =20 and 50% higher than the inception voltage and two different noise levels. The classification accuracy rate evaluated for the combination of each classifier integrated with different feature extraction algorithms are

listed in Table 1. This table presents the overall classification success rate related to the specific pairs of feature extraction/classification algorithms for each individual source of the PD.

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

As is seen in Table 1, almost all algorithms result in a desirable classification accuracy; however, FSVM, KSVM, and AdaBoost outperform the other seven algorithms. Among these classifiers, Naïve Bayes shows less accuracy that is due to the basic assumption in this algorithm which is to assume different features are statistically independent [23]. Also, SVM attains lower classification accuracy compared with KSVM and FSVM because it is a linear classification algorithm and may not be able to deal with the nonlinearity of 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 rate of 99.4 and 99.1%, respectively. FSVM and KSVM start by mapping the dataset onto a higher-dimensional feature space where in that space the classes can be classified by a hyperplane [23, 41]. The advantage of the FSVM over KSVM is that the importance of some training points can be considered in the training process. This leads to making the classifier less sensitive to the effects of noise and outliers [41]. Classifier algorithms integrated with PCA and FDA from the linear group work with higher-classification accuracy compared with statistical operators. However, classification using SPE as the feature extraction method does not show any desirable accuracy compared with other feature extraction algorithms.

7.3 Probabilistic classification

Table 1 only shows the overall classification accuracy rate of different algorithms. However, in the specific area of PD source identification, 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 studied in this study, FKNN [42], FSVM [41], and Bayesian [23, 24] have the capability to calculate the posterior probability of a test sample belonging to each class of data. Besides, as shown in Table 1, these algorithms also have a higher classification accuracy rate. To demonstrate the posterior probability calculated by these algorithms, seven data samples

Table 2 FSVM classification posterior probability rate for seven PD test samples on data output of KPCA. C: sample classified; M: misclassified; Class 1: floating electrode in SF₆; Class 2: point-plane electrodes in SF₆; Class 3: free aluminium particle in SF₆; Class 4: free aluminium particle in oil; Class 5: point-plane electrodes in oil; Class 6: point-plane electrodes in air

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

Bold values indicate the largest values.

Table 3 Bayesian classification posterior probability rate for seven 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

Bold values indicate the largest values.

Table 4 FkNN classification posterior probability rate for seven 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

Bold values indicate the largest values.

were randomly selected. The first six samples were from the samples that were correctly classified. The seventh sample was from those that were misclassified. The probabilistic classification results for these seven samples are shown in Tables 2–4 where we have used the classifiers above. In Table 2, for example, FSVM/KPCA has been employed. Each of the seven samples has a different posterior probability that shows its ‘degree of membership’ to different sources of PD. In this table, the first sample, which is originally from class 1, is determined to belong to classes 1–6 with probabilities of 84.2, 2.4, 0.4, 0.1, 12.9, and 0.0%, respectively. The seventh sample, which 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 only 30.8%.

The determination of the ‘degree of membership’ for PRPD test samples would allow safer decision making by considering the risk associated with different sources of PD in HV apparatus. The 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 from the same type of defect but does not originally belong to the original dataset. This probability shows when this sample is classified into one class, how similar this sample is to that class, and also how much is the probability of this sample belonging to other classes of data. 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 the PD for the safe operation of HV apparatus under test.

8 Conclusions

In this study, the application of an enhanced automated classification system on different sources of PD in different HV insulation media was investigated. The results are useful to increase the automated classification accuracy rate of the PD source identification in a non-time consuming way. The determination of the ‘degree of membership’ for the PRPD test samples was presented which allows safer decision making by considering the risk associated with different sources of the PD in HV apparatus. 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 associated with different PD sources into account. To collect necessary information for making a thorough dataset, laboratory experiments were performed. The laboratory measurement tests were performed on test sets that are built to model PD activities with different mechanisms in the air, oil, or SF₆. Data samples for six different classes of PD sources are captured under two levels of voltage equal to 20 and 50% higher than the inception voltage and two different noise levels. Eventually, the results of the automated classification system on insulation PD sources based on different feature extraction and classification algorithms were demonstrated.

These results show that FSVM and KSVM integrated with MDS outperform other feature extraction-classification algorithms with a classification rate of 99.4 and 99.1%, respectively. However, the application of classifier algorithms on MDS, KPCA, Isomap, SNE, and LPP show a 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. Classifier algorithms integrated with PCA and FDA from the traditional linear group show acceptable performance and they even work with higher-classification accuracy compared with statistical operators.

The 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 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.

9 Acknowledgment

The authors are very thankful to Dr M. Jafari-Jozani from the Department of Statistics, University of Manitoba, for his valuable comments.

10 References

- [1] Fothergill, J.C.: 'Ageing, space charge and nanodielectrics: ten things we don't know about dielectrics'. 2007 Int. Conf. on Solid Dielectrics (ICSD), Winchester, UK, 2007, pp. 1–10
- [2] Arman, A., Starr, A.: 'The measurement of discharges in dielectrics', *J. Inst. Electr. Eng.*, 1936, **79**, (475), pp. 67–81
- [3] Robinson, D.M.: '*Dielectric phenomena in high voltage cables*' (Chapman and Hall, London, 1936)
- [4] James, R.E., Phung, B.T.: 'Development of computer-based measurements and their application to PD pattern analysis', *IEEE Trans. Dielectr. Electr. Insul.*, 1995, **2**, (5), pp. 838–856
- [5] Sahoo, N.C., Salama, M.M.A., Bartnikas, R.: 'Trends in partial discharge pattern classification: a survey', *IEEE Trans. Dielectr. Electr. Insul.*, 2005, **12**, (2), pp. 248–264
- [6] Krivda, A.: 'Automated recognition of partial discharges', *IEEE Trans. Dielectr. Electr. Insul.*, 1995, **2**, (5), pp. 796–821
- [7] Kranz, H.-G., Krump, R.: 'Partial discharge diagnosis using statistical optimization on a PC-based system', *IEEE Trans. Electr. Insul.*, 1992, **27**, (1), pp. 93–98
- [8] Salama, M.M.A., Bartnikas, R.: 'Determination of neural-network topology for partial discharge pulse pattern recognition', *IEEE Trans. Neural Netw.*, 2002, **13**, (2), pp. 446–456
- [9] Petrov, L., Lewin, P., Czaszejko, T.: 'On the applicability of nonlinear time series methods for partial discharge analysis', *IEEE Trans. Dielectr. Electr. Insul.*, 2014, **21**, (1), pp. 284–293
- [10] Cachin, C., Wiesmann, H.J.: 'PD recognition with knowledge-based preprocessing and neural networks', *IEEE Trans. Dielectr. Electr. Insul.*, 1995, **2**, (4), pp. 578–589
- [11] Gulski, E., Kreuger, F.H.: 'Computer-aided recognition of discharge sources', *IEEE Trans. Electr. Insul.*, 1992, **27**, (1), pp. 82–92
- [12] Lalitha, E.M., Satish, L.: 'Wavelet analysis for classification of multi-source PD patterns', *IEEE Trans. Dielectr. Electr. Insul.*, 2000, **7**, (1), pp. 40–47
- [13] Wang, M.-H.: 'Partial discharge pattern recognition of current transformers using an ENN', *IEEE Trans. Power Deliv.*, 2005, **20**, (3), pp. 1984–1990
- [14] Contin, A., Cavallini, A., Montanari, G.C., et al.: 'Digital detection and fuzzy classification of partial discharge signals', *IEEE Trans. Dielectr. Electr. Insul.*, 2002, **9**, (3), pp. 335–348
- [15] Chan, J.C., Saha, T.K., Ekanayake, C.: 'Pattern recognition techniques and their applications for automatic classification of artificial partial discharge sources', *IEEE Trans. Dielectr. Electr. Insul.*, 2013, **20**, (2), pp. 468–478
- [16] Sharkawy, R.M., Mangoubi, R.S., Abdel-Galil, T.K., et al.: 'SVM classification of contaminating particles in liquid dielectrics using higher order statistics of electrical and acoustic PD measurements', *IEEE Trans. Dielectr. Electr. Insul.*, 2007, **14**, (3), pp. 669–678
- [17] Satish, L., Gururaj, B.: 'Use of hidden Markov models for partial discharge pattern classification', *IEEE Trans. Electr. Insul.*, 1993, **28**, (2), pp. 172–182
- [18] Zhang, X., Xiao, S., Shu, N., et al.: 'GIS partial discharge pattern recognition based on the chaos theory', *IEEE Trans. Dielectr. Electr. Insul.*, 2014, **21**, (2), pp. 783–790
- [19] Zhang, X., Ren, J., Tang, J., et al.: 'Kernel statistical uncorrelated optimum discriminant vectors algorithm for GIS PD recognition', *IEEE Trans. Dielectr. Electr. Insul.*, 2009, **16**, (1), pp. 206–213
- [20] Contin, A., Rabach, G.: 'PD analysis of rotating ac machines', *IEEE Trans. Electr. Insul.*, 1993, **28**, (6), pp. 1033–1042
- [21] Abdel-Galil, T., Sharkawy, R., Salama, M., et al.: 'Partial discharge pulse pattern recognition using an inductive inference algorithm', *IEEE Trans. Dielectr. Electr. Insul.*, 2005, **12**, (2), pp. 320–327
- [22] Peng, X., Zhou, C., Hepburn, D., et al.: 'Application of K-means method to pattern recognition in on-line cable partial discharge monitoring', *IEEE Trans. Dielectr. Electr. Insul.*, 2013, **20**, (3), pp. 754–761
- [23] Theodoridis, K.K.S.: '*Pattern recognition*' (Elsevier Inc., Burlington, MA, 2009, 4th edn.)
- [24] Duda, R., Hart, P., Stork, D.: '*Pattern classification*' (John Wiley & Sons, New York, 2001, 2nd edn.)
- [25] Jain, A., Duin, R.P.W., Mao, J.: 'Statistical pattern recognition: a review', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, **22**, (1), pp. 4–37
- [26] IEC 60270: 'High-voltage test techniques – partial discharge measurements', Tech. Rep., 2001
- [27] Janani, H., Kordi, B., Jozani, M.J.: 'Classification of simultaneous multiple partial discharge sources based on probabilistic interpretation using a two-step logistic regression algorithm', *IEEE Trans. Dielectr. Electr. Insul.*, 2017, **24**, (1), pp. 54–65
- [28] Hampton, B.F., Meats, R.J.: 'Diagnostic measurements at UHF in gas insulated substations', *IEE Proc., Gener. Transm. Distrib.*, 1988, **135**, (2), pp. 137–144
- [29] Reid, A., Judd, M., Fouracre, R., et al.: 'Simultaneous measurement of partial discharges using IEC60270 and radio-frequency techniques', *IEEE Trans. Dielectr. Electr. Insul.*, 2011, **18**, (2), pp. 444–455
- [30] Van Der Maaten, L.J.P., Postma, E.O., Van Den Herik, H.J.: 'Dimensionality reduction: a comparative review', *J. Mach. Learn. Res.*, 2009, **10**, pp. 1–41
- [31] Fisher, R.: 'The use of multiple measurements in taxonomic problems', *Ann. Eugen.*, 1936, **7**, (2), pp. 179–188
- [32] Schölkopf, B., Smola, A., Müller, K.-R.: 'Nonlinear component analysis as a kernel eigenvalue problem', 1998, pp. 1299–1319
- [33] Baudat, G., Anouar, F.: 'Generalized discriminant analysis using a kernel approach', *Neural Comput.*, 2000, **12**, (10), pp. 2385–2404
- [34] Cox, T.F., Cox, M.A.A.: '*Multidimensional scaling*' (Chapman & Hall, London, UK, 1994)
- [35] Agrafiotis, D.K.: 'Stochastic proximity embedding', *J. Comput. Chem.*, 2003, **24**, (10), pp. 1215–1221
- [36] Tenenbaum, J.B., de Silva, V., Langford, J.C.: 'A global geometric framework for nonlinear dimensionality reduction', *Science*, 2000, **290**, (5500), pp. 2319–2323
- [37] Hinton, G.E., Roweis, S.T.: 'Stochastic neighbor embedding', *Adv. Neural. Inf. Process. Syst.*, 2002, **15**, pp. 833–840
- [38] Roweis, S.T., Saul, L.K.: 'Nonlinear dimensionality reduction by locally linear embedding', *Science*, 2000, **290**, (5500), pp. 2323–2326
- [39] He, X., Niyogi, P.: 'Locality preserving projections', *Neural Inf. Process. Syst.*, 2004, **16**, p. 153
- [40] He, X., Cai, D., Yan, S., et al.: 'Neighborhood preserving embedding'. Tenth IEEE Int. Conf. on Computer Vision (ICCV 2005), Beijing, China, October 2005, vol. 2, pp. 1208–1213
- [41] Lin, C.-F., Wang, S.-D.: 'Fuzzy support vector machines', *IEEE Trans. Neural Netw.*, 2002, **13**, (2), pp. 464–471
- [42] Keller, J., Gray, J.M., Givens, M.R.: 'A fuzzy K-nearest neighbor algorithm', *IEEE Trans. Syst. Man Cybern.*, 1985, **SMC-15**, (4), pp. 580–585
- [43] Bishop, C.: '*Neural networks for pattern recognition*', vol. **92**, (Oxford Univ. Press, New York, NY, 1995)
- [44] Specht, D.F.: 'Probabilistic neural networks', *Neural Netw.*, 1990, **3**, (1), pp. 109–118
- [45] Freund, Y., Schapire, R.: 'A decision-theoretic generalization of on-line learning and an application to boosting', *Comput. Learn. Theory*, 1995, **55**, (1), pp. 119–139