Automated Recognition of Partial Discharges

A. Krivda

High Voltage Laboratory,
Delft University of Technology,
The Netherlands

ABSTRACT

In this work an overview of automated recognition of partial discharges (PD) is given. The selection of PD patterns, extraction of relevant information for PD recognition and the structure of a data base for PD recognition are discussed. Mathematical methods useful for the design of the data base are examined. Classification methods are interpreted from a geometrical point of view. Some problems encountered in the automation of PD recognition also are addressed.

1. INTRODUCTION

THE occurrence of PD in electrical equipment had been recognized as early as the beginning of the century. As it became clear that PD has deleterious effects on insulation [1], much effort was spent on investigating this phenomenon. New measurement methods were introduced, the physics and the chemistry of PD were studied [2-6]. PD detection gradually evolved into an indispensable tool for the evaluation of modern insulating constructions [7].

If PD is found in insulating systems, then, in many cases, it is important to identify its character, *i.e.*, internal discharges, surface discharges, corona, *etc.* Such information is vital for the manufacturer, the test institute, or the user of electrical equipment. For many years recognition was performed by eye, *i.e.*, by observation of PD patterns on the power frequency ellipse on an oscilloscope screen [3, 8]. The interpretation of the patterns on the ellipse is, however, dependent on the knowledge and experience of experts. The use of computers in PD measurements [9] opened up new possibilities for automated PD recognition [10-38]. It is the purpose of this review to give an overview of efforts in the field of automated PD recognition and provide a basic bibliography relevant to problems encountered during the development of automated PD recognition systems.

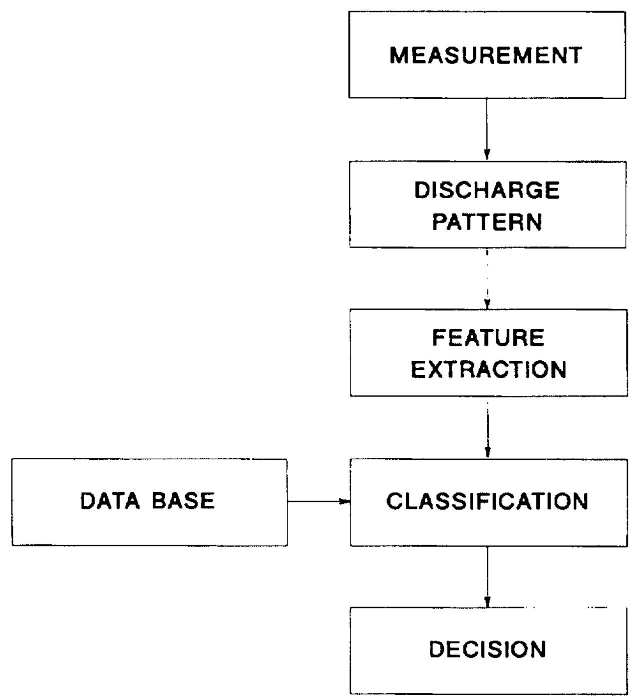

Figure 1. A general recognition procedure.

2. PD RECOGNITION

A general recognition procedure is shown in Figure 1 [39]. It consists of measurements which yield a PD pattern, feature extraction from the measured pattern, classification of the pattern, and a decision process.

1070–9878/95/$3.00 © 1995 IEEE

2.1. PATTERNS FOR PD RECOGNITION

PD measurement can be performed in many ways, *e.g.*, by measuring the charge displacement in the leads, electromagnetic waves, acoustic waves, light [3,4]. The measurement results in a pattern. There are many types of patterns which can be used for PD recognition. Those which have already been in some way employed to solve this problem are described here.

2.1.1. CHARGE DISPLACEMENT DETECTION

By measuring the charge displacement in the leads, PD pattern can be observed in the form of various discharge distributions and individual pulses.

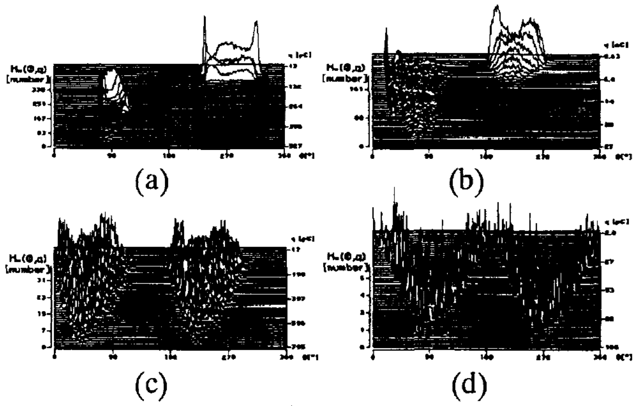

Figure 2.

$H_n(\varphi, q)$ patterns of (a) multiple-point corona in air, at the HV side, (b) surface discharges in air, with the rod at the HV side, (c) dielectric bounded cavity, (d) air bubbles in oil.

$H_n(\varphi, q)$ and related distributions

Figure 2 describes the relationship between the number, magnitude and power frequency phase angle of a PD event [40-43]. Other distributions which can be derived from the $H_n(\varphi, q)$ distribution are *e.g.*, $H_{qmax}(\varphi)$ the maximum pulse height distribution, $H_{qn}(\varphi)$ the mean pulse height distribution, $H_n(\varphi)$ the pulse count distribution, $H(q)$ the number of discharges *vs.* discharge magnitude, $H(p)$ the number of discharges *vs.* discharge energy, see Figure 3 [44-48]. The distributions can be obtained by measuring discharges with time-resolved techniques [49] or by conventional discharge detection systems (IEC 270) [48]. The distributions proved to be very useful for recognition of discharging defects [10-16, 18-26, 30, 35, 37, 38, 42, 43]. It has also been observed that the distributions can significantly change during the aging of insulation [10, 30, 41, 50-68]. This fact has been used to assess the degree of insulation degradation and the results have been encouraging [30, 53, 58, 59, 62, 64].

It should be kept in mind that many effects such as the availability of starting electrons, level of a test voltage, strongly influence the shape of the discharge distributions. These effects must be kept in mind when utilizing the distributions for PD recognition.

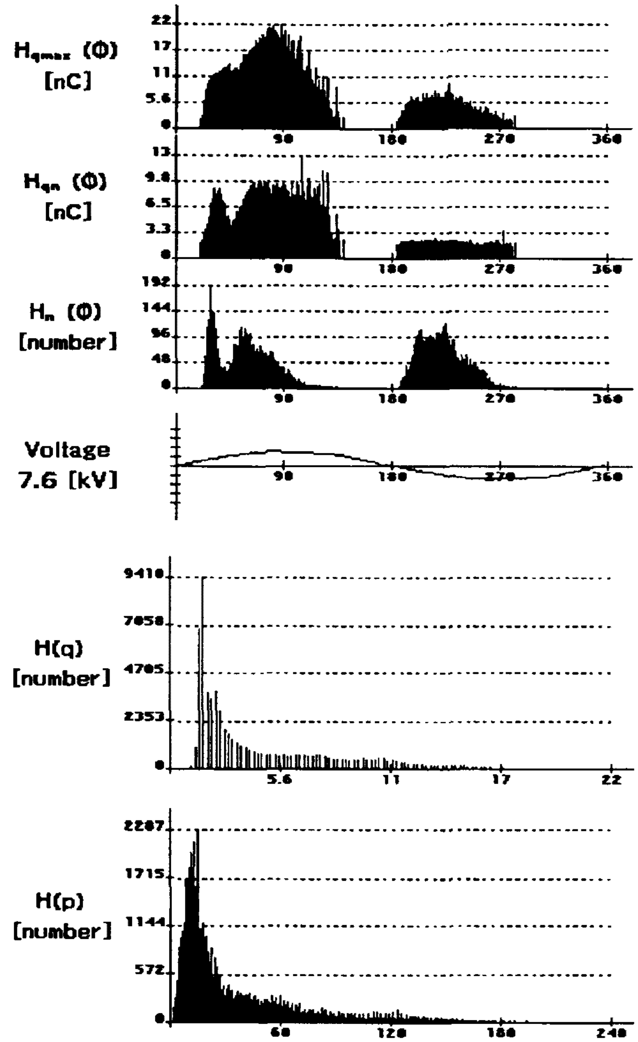

Figure 3.

$H_{qmax}(\varphi)$, $H_{qn}(\varphi)$, $H_n(\varphi)$, $H(q)$ and $H(p)$ distributions of surface discharges in air derived from the $H_n(\varphi, q)$ pattern shown in Figure 2.

Individual PD pulses

Pulses are measured with time-resolved techniques with a bandwidth of ~ 500 MHz. Distinct pulse shapes have been recorded, *e.g.*, for free conducting particles, floating parts, and treeing discharges in gas insulated systems (GIS) [27, 69, 70]. The pulse width was successfully used to discriminate between PD and noise pulses measured by stator slot couplers [71]. It has also been observed that the pulse shape of discharges originating from cavities significantly changes with time as a result of PD-induced aging [72-80]. This fact could possibly be used for the monitoring of insulation systems. Attempts to recognize the degree of insulation degradation during aging tests on the basis of the discharge pulse shape have already been undertaken [30, 34, 77]. It should be realized, however, that the PD pulse shape can exhibit great variability due to, e.g. gas composition, gas pressure, or statistical effects such as the availability of starting electrons, which influences overvoltage over a defect [81-89]. Further, propagation effects in cables, generators, etc. significantly distort the pulse shape; the loss of some frequency components, damping, reflections may occur [49, 69, 70, 90-93]. These effects must be kept in mind when designing systems for automated PD recognition.

Conditional PD distributions

These describe memory propagation effects [66, 88, 94], e.g., $p_1(\varphi_i^- \mid Q^+)d\varphi_i^-$ the probability that i-th pulse on the negative half cycle will appear between phase φ_i^- and $\varphi_i^- + d\varphi_i^-$ if the total charge associated with all discharge pulses on the previous positive half cycle equals Q^+. The use of conditional distributions for PD recognition has already been suggested [94] but their true potential has yet to be assessed.

Interpulse distributions

These derive from direct pulse-to-pulse correlations, such as the distribution of voltage or time differences between two consecutive pulses [62, 88, 95, 96]. Different distributions were recorded for different stages of the tree growth in insulation [96]. Also $H_n(q, \Delta t)$ distribution, which describes the relationship between the number, the magnitude and the time between two successive discharges has been suggested for PD recognition [97, 98]. This distribution appeared to be useful for the recognition of different discharge sources at dc voltage.

2.1.2. ELECTROMAGNETIC WAVES

By measuring electromagnetic waves PD patterns can be observed in several forms.

EM spectrum

The relative amplitude vs. frequency is measured of the electromagnetic phenomena (scanned from ~ 10 kHz to ~ 2 GHz). Defects in GIS, such as corona, free conducting particle, floating electrodes, particles on the spacer produce distinct spectra [99-103]. Glow, streamer and leader corona also resulted in distinct spectra [104]. Radio frequency spectrum of currents in the generator neutral lead as measured by a radio frequency current transformer clamped around the neutral and radio noise meters have been used successfully for the recognition of discharges within the stator windings insulation, arcing between adjacent ends of a broken coil strand, etc. [105-111].

Phase-related distributions

PD events are displayed in the form of phase related distributions, such as $H_{qmax}(\varphi)$, $H_n(\varphi)$ etc. The discharges are recorded at a particular frequency in a range of ~ 300 MHz [100, 103, 112, 113].

Interpulse distributions

These show the time difference between two consecutive pulses. The distribution was used to determine whether a free conducting particle in GIS had the capability of reaching the busbar and the associated probability to trigger breakdown [113].

2.1.3. ACOUSTIC DETECTION

By measuring acoustic waves PD patterns can be observed in the form of:

Individual pulses

Pulses have been used for the recognition of faults in GIS [17].

Frequency spectrum

This is the relative magnitude vs. frequency of the sound waves (scanned to ~ 2 MHz). Distinct spectra have been obtained for various PD sources in oil (corona, floating electrode, cavities) [114, 115] and GIS (corona, free conducting particle, particle on the spacer) [101, 116]. There is also strong evidence that the size of cavities in insulation could be estimated from the spectra [114].

An interesting approach was used in [117], where acoustic signals recorded in the time domain were transformed to Fourier space for each voltage cycle and averaged in the Fourier space for a certain period. Distinct spectra were obtained for various defects in GIS. Another possibility is to use the frequency spectrum as above, but individual pulses are related to the power frequency phase angle. This discharge distribution was employed to identify various defects in GIS [116].

$H_n(\varphi, q)$, $H_{qn}(\varphi)$, $H_n(\varphi)$ distributions

These are discussed in Section 2.1.1 but are measured acoustically [118-120]. Different distributions were obtained for branch, bush-like and filamentary trees [119].

Impact spectrum

The frequency of bouncing of a free conducting particle in GIS has been used to identify the mode of the particle, i.e., dancing, crossing or oscillating [121].

2.1.4. LIGHT

By measuring photon emission, PD patterns can be observed.

Individual PD pulses

Pulses are measured with fast photon counting systems on a ns scale [122,123]. Distinct pulses were recorded for Townsend and streamer discharges in cavities [122].

Emission spectrum

The amplitude *vs.* wavelength of emitted light [124--127] can be recorded. Defects in GIS, such as corona, floating particles, free conducting particles resulted in distinct spectra [127]. The spectrum of PD also changed during the development of carbon traces on the surface of organic materials exposed to surface discharges [125]. Corona discharges in nitrogen, helium, air and SF_6 resulted in distinct spectra [124,126].

Various phase-related distributions

These have been employed in the observation of the early stages of electrical breakdown where electroluminescence related to charge injection, the development of microchannels, and tree growth has been studied [128--137].

2.1.5. INFRARED RADIATION

Pulses recorded in the infrared part of the optical emission spectrum correlated with the ac test voltage were used to identify the formation of leader discharges in GIS [103].

Note that not all possible combinations 'measuring method/PD pattern' were mentioned. For example, conditional, direct pulse-to-pulse PD distributions can also be constructed, among others, for PD measured acoustically.

It can be seen that there are basically two types of patterns: individual pulses and various distributions (unconditional, conditional, direct pulse-to-pulse distributions, frequency spectrum, *etc.*). For recognition purposes, a type of PD pattern must be selected which is able to distinguish between various PD sources. The abovementioned patterns are, in many cases, indeed suitable for this task. This can be seen in Figure 2, where $Hn(\varphi, q)$ distributions of four PD sources are shown. Different PD sources produced different PD patterns.

The choice of detection method depends on local conditions, *e.g.* when electrically based measurement is impossible due to the high level of disturbances, an acoustic measurement of PD to measure the patterns can be considered. It should be borne in mind that the measurement must be executed carefully and use must be made of all available knowledge in the field [3,4,90,138-140]. Badly performed measurements cannot provide a reliable basis for PD recognition.

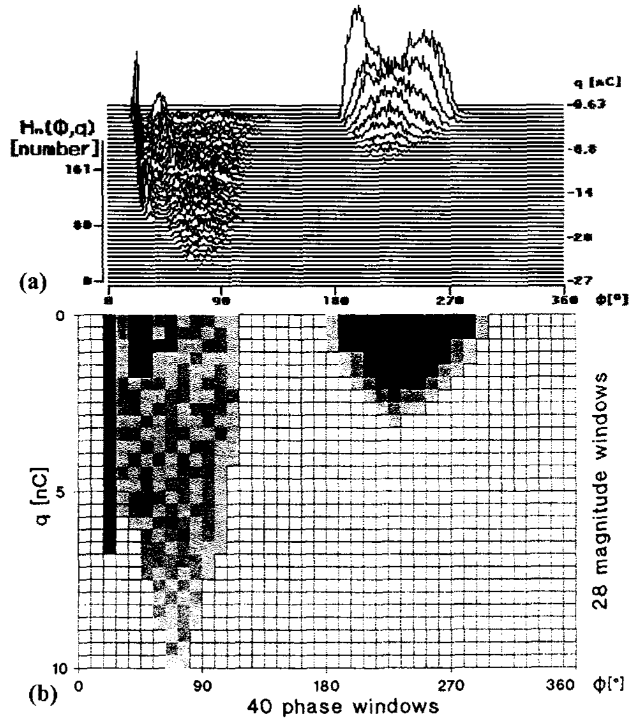

Figure 4.

(a) $H_n(\varphi, q)$ pattern of surface discharges in air, with the rod at the HV side, in a 3-d view. (b) Two-dimensional projection of the $H_n(\varphi, q)$ pattern. Gray level indicates the number of PD.

2.2. FEATURE EXTRACTION FOR PD RECOGNITION

The following step in automated PD recognition is the feature extraction from measured discharge patterns. The aim of feature extraction is to reduce the dimensionality of an original PD pattern by calculating certain'features' or'properties' of the pattern [39,141-144]. The essential condition here is that the features must distinguish between different PD sources as well as original patterns do. The features are usually derived on the basis of past experience, theoretical insight, intuition or simple guesswork [145]. The number of features should be as low as possible. The lower the number of features, the faster the speed of classification. Also, the features must be extracted from original patterns in real time, *e.g.* in 1 s on moderately fast computers. Consider, for example, the $H_n(\varphi, q)$ pattern shown in Figure 4. Even with low resolution, say, 40 phase windows and 28 magnitude windows, a total of $40 \times 28 = 1120$ parameters have to be employed for recognition. This is not a very efficient way of handling the information due to memory consumed, computational complexity, and so on. By describing the $H_{qmax}(\varphi)$, $H_{qn}(\varphi)$, $H_n(\varphi)$, $H(q)$ and $H(p)$ distributions (which can be derived from the basic $H_n(\varphi, q)$ pattern), see Figure 3, by statistical parameters such as skewness,

Table 1.
Resulting values of 29 statistical parameters, features, describing the shapes of the distributions in Figure 3.

Feature	Distribution $H_{qmax}(\varphi)$	$H_{qn}(\varphi)$	$H_n(\varphi)$	$H(q)$	$H(p)$
Skewness+	−0.04	−0.04	0.18	1.24	1.75
Skewness-	0.27	0.24	0.34		
Kurtosis+	−0.88	−0.93	−0.94	0.70	2.65
Kurtosis-	−0.72	−0.80	−0.64		
Peaks+	4.00	3.00	2.00		
Peaks-	2.00	3.00	3.00		
Asymmetry	−0.55	−0.53	−0.02		
cc	0.50	0.49	0.36		
Phase		14.30			

kurtosis and asymmetry, a total of 29 features was calculated and successfully used for PD recognition [25,48]. In Table 1, an example of a set of 29 statistical parameters, called a 'fingerprint', is shown. The computation of all 29 features took ~ 3 s on a 486 computer. Note that a significant reduction in recognition parameters was achieved: 29 instead of 1120 parameters.

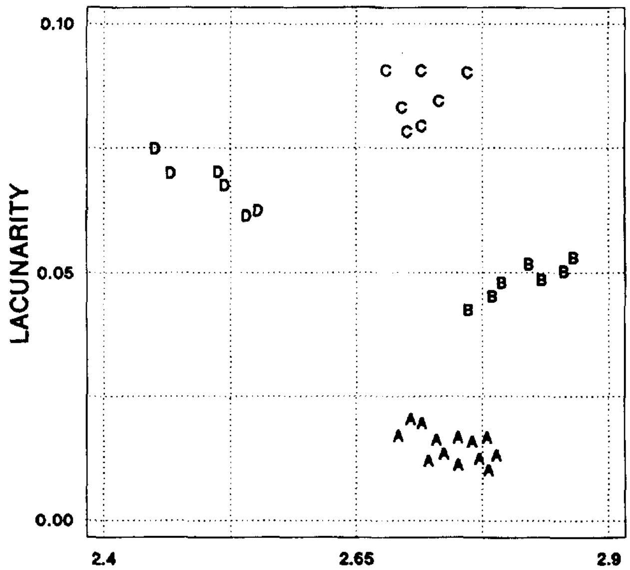

Figure 5.

Fractal dimension and lacunarity calculated for the $H_n(\varphi, q)$ patterns shown in Figure 2. Each letter represents a single $H_n(\varphi, q)$ pattern: A: multiple-point corona in air, at the HV side, B: surface discharges in air, with the rod at the HV side, C: dielectric bounded cavity, D: air bubbles in oil.

Recently, fractal features have been introduced to describe $H_n(\varphi, q)$ patterns [35]. In this case, a $H_n(\varphi, q)$ pattern was reduced to just two dimensions by calculating fractal dimension and lacunarity from the pattern. Fractal dimension describes surface roughness, and lacunarity the denseness of the $H_n(\varphi, q)$ pattern, features which are apparently relevant descriptors of the $H_n(\varphi, q)$ patterns. Figure 5 shows fractal features calculated for the $H_n(\varphi, q)$ discharge patterns shown in Figure 2. It can be seen that two fractal features enable distinction between PD sources as well as the original 1120 dimensional $H_n(\varphi, q)$ pattern can. This can be considered to be a very efficient feature extraction.

Another example of feature extraction is the description of a PD pulse shape as measured with time-resolved techniques by its peak amplitude, area of the pulse, rise time, fall time, and width [34]. In this case point-by-point digital description of discharge pulse (e.g. by 256 points) was replaced by five parameters. Another possibility would be to describe the PD pulse shape by several Fourier transform [146] or wavelet transform [147] or Karhunen-Loève transform [148] coefficients. In all cases a significant reduction of the dimensionality of original PD data was achieved.

In addition to the 'quantitative' features discussed above, also 'qualitative' or 'abstract' features can be used for recognition. Such abstract features have been used in the development of an expert system for fault recognition in GIS [17] in the form of, e.g. the physics of acoustic wave propagation, the knowledge about the architecture, and the operation of GIS.

It should be realized that the features are usually pure descriptors of original PD patterns, and they are not necessarily predictable in terms of basic physical processes [88]. But if the aim is discharge recognition, the use of such features for recognition purposes can be justified.

When features are calculated, it is important to verify whether the features indeed contain sufficient information that can be used for PD recognition. When only two features are calculated, then such a check can easily be performed by making a scatter plot of data in feature space as has been done in case of fractal features, see Figure 5. However, there are usually far more features calculated, *e.g.* 29 statistical parameters for the description of $H_{qmax}(\varphi)$, $H_{qn}(\varphi)$ distributions, five features for the description of the PD pulse shape in the example above, and so on. If this is the case, the use of rather complex mathematical techniques for discriminating among PD patterns is unavoidable [39, 141, 144, 149-154]. Some of the techniques are briefly discussed in this work.

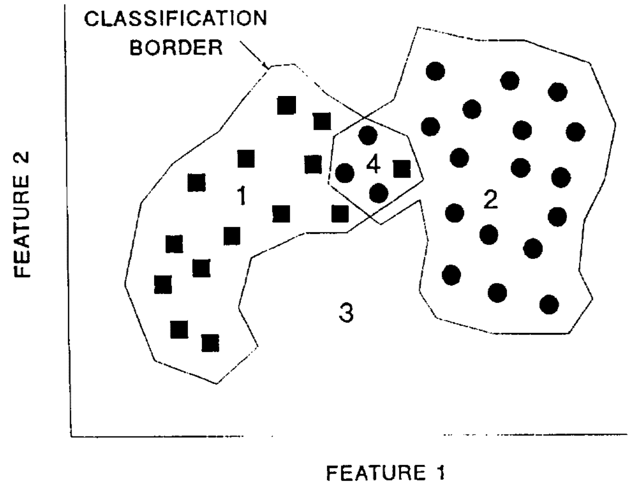

Figure 6.

Four possible cases of classification of a pattern of unknown origin in feature space. ■: pattern of defect A, ●: pattern of defect B, 1: pattern classified to defect A, 2: pattern classified to defect B, 3: pattern classified to none of the defects, 4: pattern classified to both defects. A single point represents one PD pattern.

2.3. CLASSIFICATION AND DECISION

The aim of classification is to assign a label to a PD pattern of unknown origin from previously collected patterns with known labels, such as treeing discharges, corona, *etc.* This means that a data base of previously collected patterns must be available, and that feature space must be partitioned into regions, *i.e.*, borders between patterns of various PD sources must be defined. In the case of two defects, say defects A and B, four possible situations of classification are possible, see Figure 6. A pattern of unknown origin can belong to defect A, defect B, none of them, and in the case of an overlap between defects A and B, to both defects. The last case can occur when features are badly designed or when there is indeed no difference between patterns from which features were extracted.

There are a number of methods available for classification purposes: conventional classifiers, fuzzy classifiers, neural networks [39, 141-143, 154-173]. Some of the methods and the way they create borders for classification purposes are discussed below.

On the basis of the classification result, *i.e.*, when a PD pattern of unknown origin has been identified such as treeing discharges, a decision has to be made, for instance 'go', 'no-go' decision for the operation of power apparatus [174]. The decision is based on the knowledge of the potential danger of different defects derived from past experience. The decision is usually made by humans, although the decision process also can be automated [175-177].

3. DATA BASE FOR PD RECOGNITION

Provided that a type of PD pattern has been selected that is able to distinguish between various PD sources, and features with a sufficient discriminating power have been calculated from the pattern, then a data base of previously collected patterns can be created. Patterns of unknown origin can now be compared to the known patterns stored in the data base. A carefully designed database for PD recognition should produce the high level of similarity between a fingerprint to be classified and an insulation defect in the case of a correct recognition, and the low level of similarity in all other cases. Several questions arise when creating the data base. For example, what should the structure of the data base be? How many features are sufficient for recognition? Are stored patterns in the data base representative of a particular PD source? How many patterns of one and the same PD source are required for future successful classification?

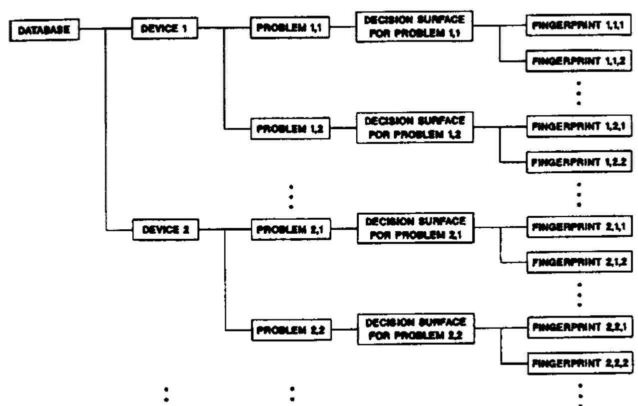

Figure 7.

The structure of a data base for PD recognition.

3.1. STRUCTURE OF A DATA BASE

A reasonable way of constructing of a data base for PD recognition is to divide the data base into several levels [48]. An example of such a data base structure is shown in Figure 7. It can be seen that this data base has four levels. The first level consists of ***fingerprints***, which are the basic elements used for PD recognition. An example of a fingerprint can be a set of 29 statistical parameters describing the shapes of $H_{qmax}(\varphi)$, $H_{qn}(\varphi)$, *etc.*, distributions as has been discussed above.

The second level consists of ***decision surfaces*** or rules or functions derived from the fingerprints. An example of such a decision surface is a classification border shown in Figure 6. After extraction of the decision surfaces one does not use the fingerprints for the classification but the decision surfaces. The formation of decision surfaces from fingerprints is discussed in detail in Section 4.

The third level consists of ***problems***. A problem is a label (name) of a particular PD source. Examples of problems are: 'Dielectric bounded cavity' which consists of decision surfaces derived from, *e.g.* ten fingerprints of ten different models containing a dielectric bounded cavity. The problem 'Corona in air' consists of decision surfaces constructed from, *e.g.* eight fingerprints of eight identical setups causing corona in air. A fingerprint of unknown origin can now be assigned to the reference problems by using decision surfaces of the problems. The result can then be expressed as the percentage of resemblance of the fingerprint of unknown origin to the known problems in the data base.

The term 'problem' originates from practice. For example, if a manufacturer encounters three types of defects in a production process, the fingerprints of these defects can be stored in three, presumably different, problems in the data base.

The fourth level, called *device*, is a collection of several problems. Examples of devices might be 'GIS', 'High Voltage transformer' among others. The device GIS might then consist of problems such as 'Corona in SF_6', 'Free conducting particle', or 'Particle on the spacer'.

The main reason for such a data base configuration originates in industry. For instance, a manufacturer of GIS might consider irrelevant problems in large power transformers such as surface discharges in oil and corona in oil. Further, it is convenient to make comparisons of a fingerprint of unknown origin with as few as possible known problems. First, the fewer the number of problems, the better the recognition: if the number of problems decreases, so does the chance of an overlap between the problems. Second, it saves computation time.

3.2. QUALITY OF A DATA BASE

To create a quality data base for PD recognition, attention must be paid to several issues.

(1) Effects influencing discharge patterns should be taken into account. For example, for phase-related distributions, changes in the level of a test voltage, the availability of starting electrons [88,43]. Also, as a result of PD induced aging, patterns can change with time [30,41, 50,53,58,59,62-68]. To build a data base where PD patterns with all possible variations in voltage levels, availability of electrons, aging time, *etc.*, could be stored, is a very difficult task. Some progress has been made in recent years by creating data bases with collected patterns, *e.g.* at 1.5 discharge inception voltage or with effects of aging taken into account [10-38], but much work has yet to be done.

(2) The number of features used for recognition must be determined [39]. Usually as many features as possible are extracted from patterns [154]. This is done in the expectation that the features will discriminate not only PD sources collected at the present time but also in the future. Also the cost of experiments can be high.

However, the higher the number of features, the longer the time required to calculate the features, and the longer the classification takes. Some of the features also may be useless for pattern recognition purposes, because they may have no discriminating power. An optimum between these two opposite objectives must be found. If it appears that a reduction in the number of features is necessary then forward selection strategy could be used [154]. The method selects the best m out of p features by maximizing (or minimizing) certain criteria. In this case, it could be called 'recognition power of features'. The method first examines all features to find the one which is the best for recognition. This single 'best' feature is retained and tried in conjunction with each of the remaining $(p-1)$ features to find the pair which is the best for recognition. This pair is retained and tried with $(p-2)$ features to find the triple which is the best for recognition, and so on. It takes a few minutes on a moderately fast computer to select a few best features out of ten when $\sim$ 500 patterns are involved in the procedure. Other approaches, such as dynamic programming and a branch and bound algorithm can be found in [178-180]. It should be realized, however, that even with these methods selected features may not be optimal and only an exhaustive search over all possible combinations of features can give a best subset of features [181].

(3) The number of fingerprints in each problem should be approximately 5 to 20$\times$ the number of features [160]. This is known as the 'curse of dimensionality' [39,145]: the higher the number of features, the higher must be the number of fingerprints to fill the multidimensional space, and to determine correctly the borders between patterns of PD sources in the feature space. This is illustrated in Figure 8, where the true and the small sample size population of two defects is shown in the feature space. A data base with only a small number of fingerprints is insufficient for the correct determination of borders between defects, no matter whether neural networks, the

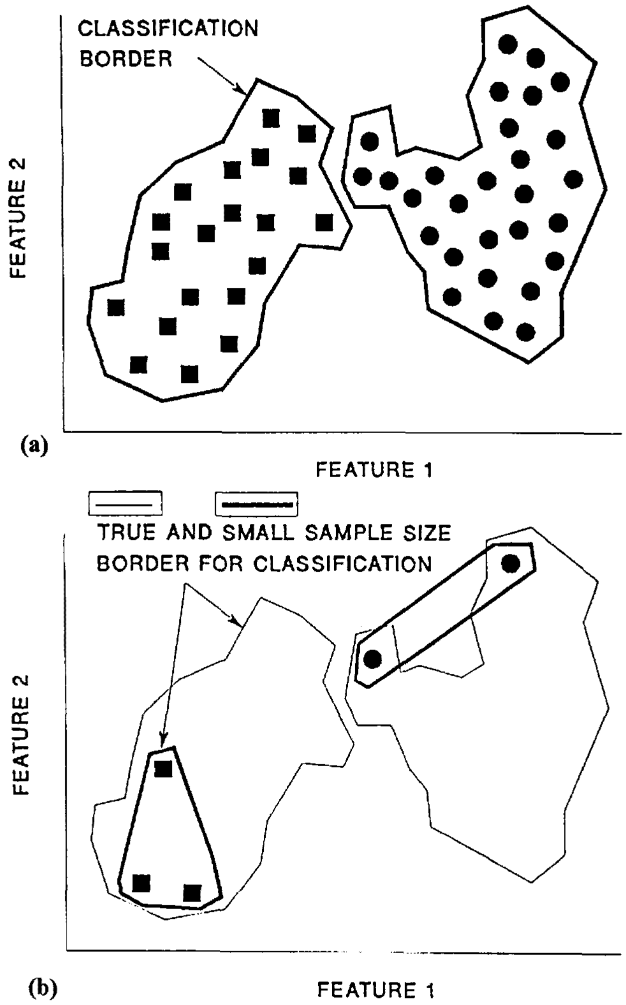

Figure 8.

Classification borders for (a) true and (b) small sample size population of fingerprints of two PD sources in feature space. A single point represents one PD pattern. ■: pattern of defect A, ●: pattern of defect B.

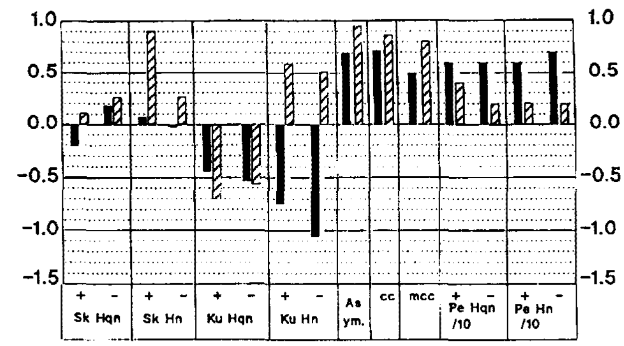

Figure 9.

A visual comparison of two fingerprints: cavity and treeing discharges in 6/10 kV polyethylene cable. The fingerprints consist of 15 values of statistical parameters (the skewness, the kurtosis) describing the shapes of $H_{qn}(\varphi)$ and $H_n(\varphi)$ discharge distributions. Data were taken from [15].

centour score, or some other classification method is used for a further classification of a fingerprint of unknown origin [182-184]. The relationship between the number of features and the number of fingerprints is, however, not linear but exponential. The heuristic rule mentioned above is in many cases sufficient. It is also important to have a large number of fingerprints in order to describe statistical effects in measured PD phenomena. For example, patterns of the same PD source measured under identical conditions, *i.e.*, the same test voltage level, the same test sample, can be highly similar but are rarely the same. This will be reflected in the scatter of feature values extracted from the patterns.

(4) It should be verified whether calculated features such as statistical parameters, discriminate between PD

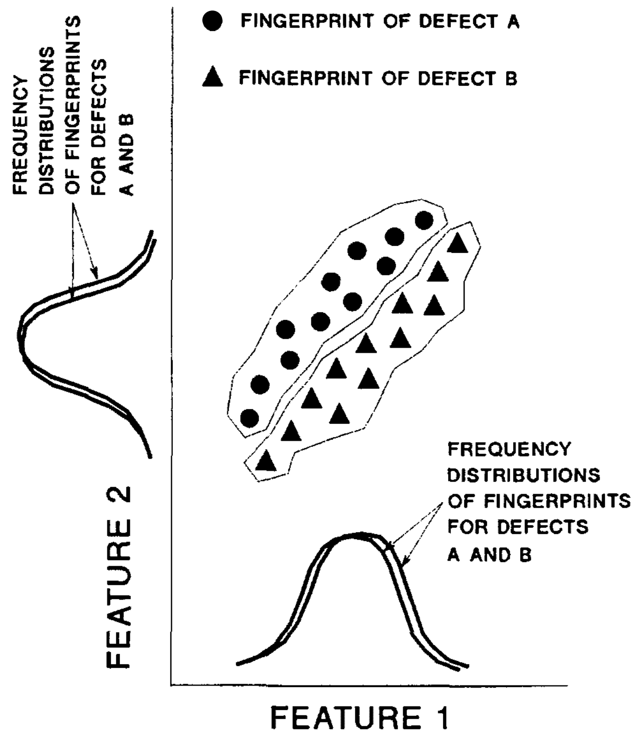

Figure 10.

Fingerprints of two different defects in feature space with mutually correlated features. A single point represents one PD pattern. No difference between two groups can be found when the groups are examined feature by feature.

sources as well as original PD patterns, *e.g.* $H_{qmax}(\varphi)$, and $H_{qn}(\varphi)$, *etc.* distributions [39, 141, 144, 154]. When only two features are extracted from a discharge pattern, the scatter plot of data in feature space provides sufficient information on the discriminating abilities of the features, see Figure 5, where the scatter plot of data in fractal feature space is shown. Usually there are far more than two or three features extracted from original discharge patterns, *e.g.* 29 statistical parameters which describe the shapes of $H_{qmax}(\varphi)$, $H_{qn}(\varphi)$, *etc.*, distributions. In this case, a simple method to assess differences (or the similarity) between fingerprints would be to look at individual features as shown in Figure 9. This method, however, does not reveal differences between fingerprints when features are mutually correlated, see Figure 10. By looking at each feature separately, no difference between two groups would be found. To overcome this problem, the use of mathematical techniques which reduce multidimensional feature space to two or three dimensions is recommended [39, 141, 144, 154]. Fingerprints of PD sources are then viewed in this new, two or three-dimensional space and the structure of data is assessed. The knowledge on the data structure in the feature space can also help in the selection of a classification method. Two groups of methods can be used to discover structures in data: mapping techniques and cluster analysis methods [39, 141, 144, 149-154, 185-193].

3.3. MAPPING TECHNIQUES

Mapping techniques project fingerprints in a multidimensional feature space on new artificially created axes. The number of new axes is substantially lower than the dimension of the original space. More than fifteen methods are available for this purpose [194-206]. Two of the methods, the principal component analysis and discriminant analysis which are widely used and can be found in most commercial statistical packages, are briefly discussed.

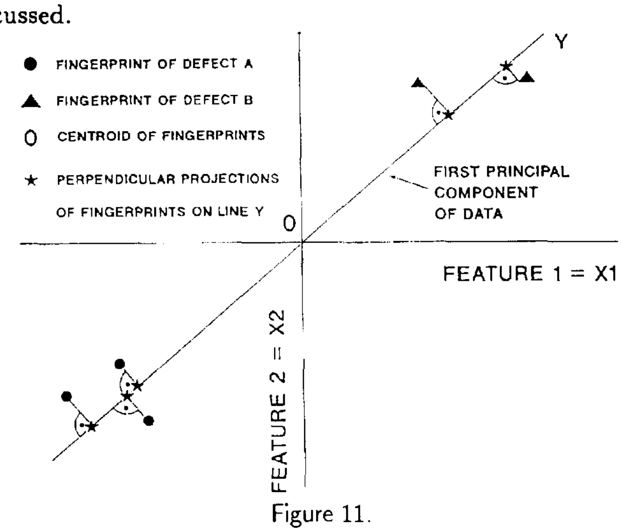

Figure 11.
Principal component analysis of data.

3.3.1. PRINCIPAL COMPONENT ANALYSIS

This mapping technique reduces the dimensions by finding principal components of the data. Principal components are linear combinations of features which describe the maximum variance in the data (the maximum spread in the data) [39, 144, 154].

To understand the principle of this method, five fingerprints in a two-dimensional space are shown in Figure 11. It can be seen that perpendicular projections of the fingerprints onto *one line* Y, give positions of the fingerprints with almost the same efficiency as that obtained with two features. The line Y which is the principal component of the data, is in this example given by a linear combination of the two features

$$Y = a_1X_1 + a_2X_2 \tag{1}$$

where X_1 and X_2 are the first and the second feature. The parameters a_1 and a_2 are determined by maximizing the *variance* in the data. No *a priori* knowledge of the membership of individual fingerprints to a particular defect is required in this analysis.

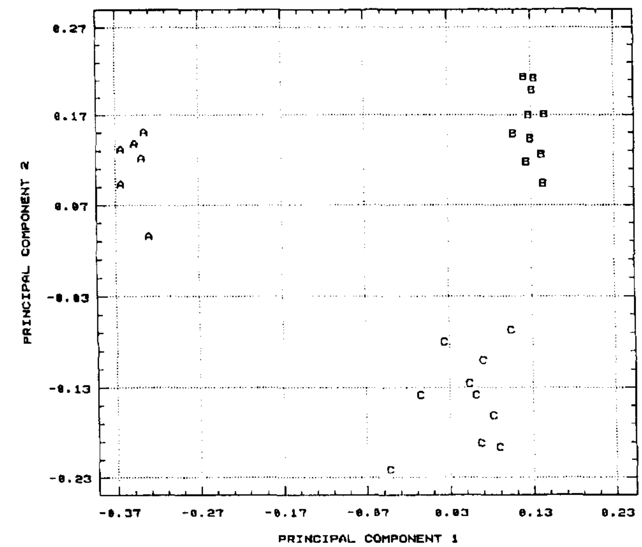

Figure 12.

Resulting scatter plot of fingerprints on two principal components. The original 29-dimensional space of statistical parameters was reduced to a 2-dimensional space of the principal components. Each letter represents a single fingerprint. A: single-point corona in air at the HV side, B: dielectric bounded cavity and C: surface discharges in air with the rod at the HV side.

An example of the resulting plot of actual fingerprints onto two principal components is shown in Figure 12. Here, the 29-dimensional space of statistical parameters (skewness, kurtosis, *etc.*, describing the shapes $H_{qmax}(\varphi)$, $H_{qn}(\varphi)$, *etc.*, distributions as shown in Figure 3) was reduced to a two-dimensional space of principal components. Each letter represents a single fingerprint. Three main groups can be identified: A stands for fingerprints of single-point corona in air at the HV side, B corresponds to a dielectric bounded cavity and C indicates surface discharges in air with a rod at the HV side. It can be seen that the first component separates corona discharges from the other defects. The second component separates the dielectric bounded cavities from surface discharges. The separation of groups was successful in this case. It should be also noted that 29 statistical parameters (in this case their linear combination) indeed contain sufficient information to distinguish between various discharge sources.

The principal component analysis provides a quick way of assessing the structures in data. However, it should be kept in mind that maximizing the *variance* in the data may not necessarily reveal actual groups in the data, as is shown in the following section on discriminant analysis.

3.3.2. DISCRIMINANT ANALYSIS

In discriminant analysis, a linear combination of p features $X_1, X_2, \ldots, X_p$

$$D = u_1X_1 + u_2X_2 + \cdots + u_pX_p \tag{2}$$

is called the discriminant function [194]. The factors u' are weights determined by maximizing the ***separation*** between defects. To perform the analysis, ***a priori*** knowledge of the membership of individual fingerprints to a particular defect is required.

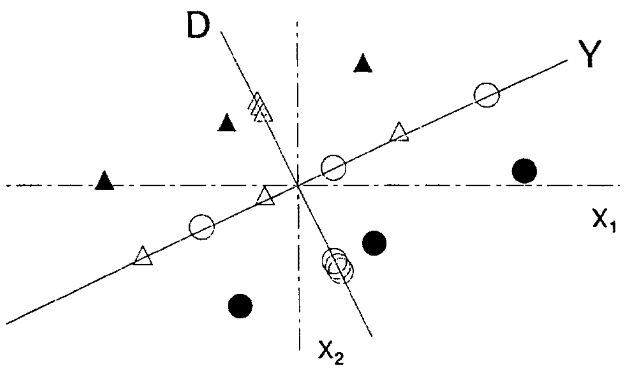

Figure 13.

Example of discriminant analysis and principal component analysis applied to two-dimensional data. X_1 and X_2 are features, the line D is the discriminant function and the line Y is the principal component of the data. ●: fingerprint of defect A, ▲: fingerprint of defect B, O: perpendicular projection of defect A fingerprint, △: perpendicular projection of defect B fingerprint,

The difference between discriminant analysis and principal component analysis is illustrated in Figure 13, where two elongated clusters are shown. Discriminant analysis creates the maximum separation between the clusters, which is found at line D. Perpendicular projections of the fingerprints on this line clearly distinguish the clusters. Principal component analysis maximizes the variance in the data. In this example, perpendicular projections on line Y, which is the principal component of the data, do not reveal differences between the clusters.

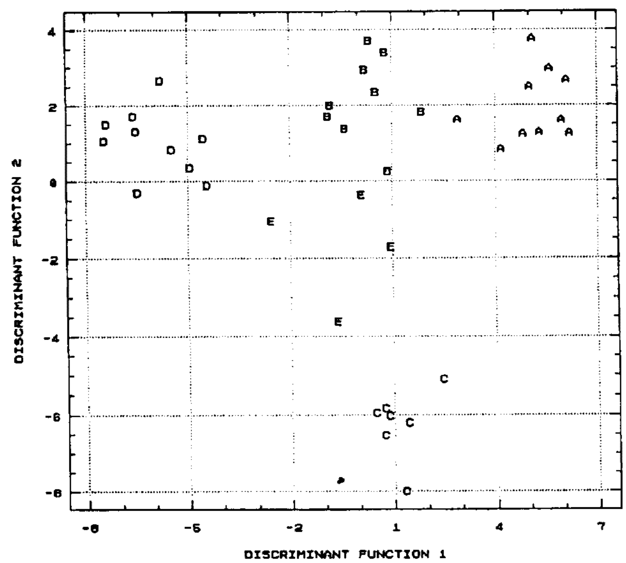

Figure 14.

Discriminant analysis applied to actual PD data. The original 15-dimensional space of statistical parameters (the skewness, the kurtosis, *etc.* describing the shapes of $H_{qn}(\varphi)$ and $H_n(\varphi)$ distributions) was reduced to a 2-dimensional space of discriminant functions. Each letter represents a single fingerprint. A: dielectric bounded cavity, B: surface discharges in air, C: surface discharges in oil, D: corona in air and E: floating parts in air. The analysis was performed on data taken from [15].

An example of discriminant analysis applied to actual PD data is shown in Figure 14. The fingerprints in the 15-dimensional space of statistical parameters (skewness, kurtosis, *etc.* describing the shapes of $H_{qn}(\varphi)$ and $H_n(\varphi)$ distributions [15]) were mapped to two dimensions. The two discriminant functions successfully separate the fingerprints of five different defects.

The discriminant analysis yields the best linear separation between defects in a multidimensional space. When compared to the principal component analysis, the discriminant analysis requires *a priori* knowledge of the membership of individual fingerprints to a particular defect.

Principal component analysis and discriminant analysis are linear methods, *i.e.* they separate data in a linear fashion. To separate data in a nonlinear way, other methods have to be used, such as multidimensional scaling [195,196], Sammon's nonlinear mapping [197], *etc.* [154,205]. The use of mapping techniques is restricted to a small number of fingerprints (usually < 200). As the number of fingerprints increases, scatter plots of data might become rather unclear.

3.4. CLUSTER ANALYSIS

Cluster analysis also tries to recognize groups without a priori knowledge. This means that no labels indicating the membership of an individual fingerprint to a particular defect are required. There are more than a hundred algorithms available for the clustering of data (hard and fuzzy partition algorithms) and it should be realized that there is no 'best' procedure [39,141,144,149,150, 154,185-193,207-216]. The use of cluster analysis is illustrated with the group average clustering method. The method forms final groups in data in the following way [39,144,-154,217].

1. Each fingerprint is declared as a group and distances between all groups are calculated;
2. Two groups with the smallest distance are fused together and declared to be one group. In this way the total number of groups in the data is reduced by one.
3. Distances between all groups are again calculated. The choice of distance is important. The group average method calculates the average distance between two groups.
4. Steps 2 and 3 are repeated until just one group is left.

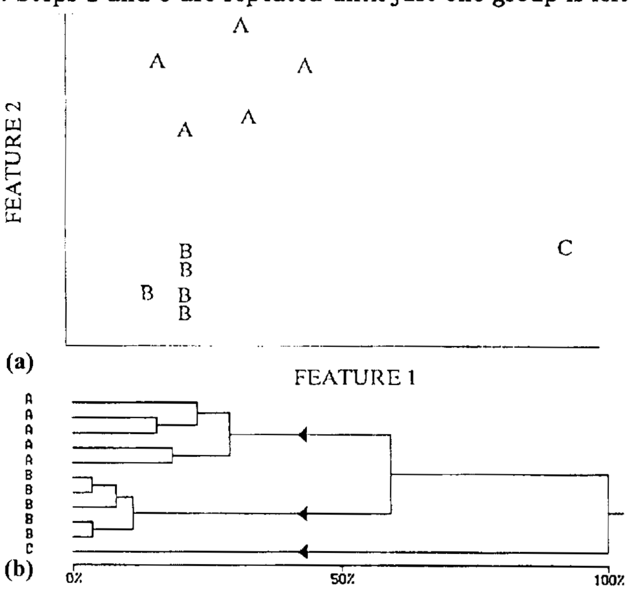

Figure 15.

(a) Scatter plot of fingerprints in feature space. A single letter represents one PD pattern. (b) Tree structure of the data in (a) obtained by the group average clustering method.

The algorithm results in a tree structure which allows a detailed examination of the relationship between individual fingerprints. The use of the method is explained in an example shown in Figure 15, where the scatter plot of fingerprints in feature space and the corresponding tree structure is shown. The scale in the lower part of Figure 15(b) shows the dissimilarity between fingerprints as a percentage of the distance between the last two groups that were fused together. In this example the last two groups were AB and C. It follows that similar fingerprints will be connected at relatively low dissimilarity levels, whilst differing fingerprints will be connected at relatively high dissimilarity levels. By looking for main 'branches' in such tree structures, different groups of fingerprints can be identified. It can be seen that a fingerprint C is an outlier in the data and does not belong to any of the known groups. This is reflected in the tree structure where a separate branch for the fingerprint was formed. Similarly, separate branches for fingerprints in the groups A and B were formed. Because of the larger spread of fingerprints in group A in the feature space, the individual fingerprints of this group are mutually connected at higher dissimilarity levels in the tree structure.

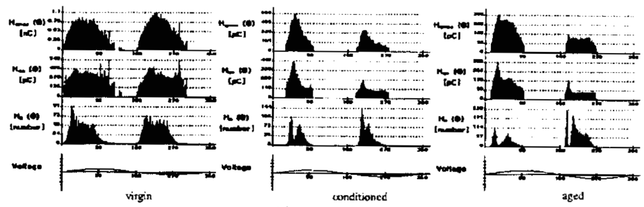

Figure 16.

$H_{qmax}(\varphi)$, $H_{qn}(\varphi)$ and $H_n(\varphi)$ distributions of electrode bounded cavity at the HV side obtained during first 2 min (virgin stage), after 5 min (conditioned stage) and after 90 min (aged stage) of the voltage application. Distributions were collected over a period of 2 min at a voltage level 50% above discharge inception.

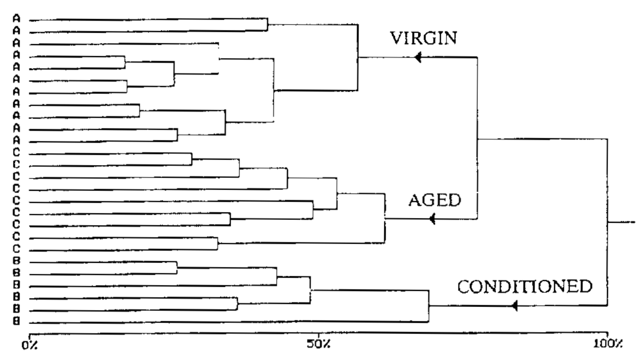

Figure 17.

Tree structure for fingerprints of electrode bounded cavities collected at different times after the voltage application. Each letter represent a single fingerprint. A: cavity aged 2 min (virgin stage), B: cavity aged 5 min (conditioned stage), C: cavity aged 90 min (aged stage). Corresponding discharge distributions are shown in Figure 16.

To demonstrate the use of the method on actual PD data, fingerprints collected during 90 min aging of electrode bounded cavities (diameter 5 to 9 mm, height 0.4 to 0.5 mm) were analyzed [67]. PD distributions significantly changed during this relatively short period, see Figure 16 and Table 2, where the distributions obtained during first 2 min, after 5 min and after 90 min are shown. The distributions were collected over a period of 2 min at a voltage level 50% above the discharge inception (PD data were collected from at least seven samples of cavities per aging time). From the distributions, 29 statistical parameters were calculated and analyzed by the group average method, without any a priori knowledge of the aging time when the fingerprints were collected. The tree structure obtained by the method, see Figure 17, reflects changes in the PD distributions and groups the data according to their aging time. On the basis of such a tree structure a data base of PD patterns obtained during aging can be made and used for recognition in the future. The method was extensively applied to analyze PD data obtained during aging on a number of artificial defects and industrial HV components, with encouraging results [218-220].

Table 2.

Classification of fingerprints of each aging stage from Figure 16 by the centour score method. Each classification category was represented by at least 210 fingerprints. HVelb: HV electrode bounded.

	HVelb virgin	HVelb conditioned	HVelb aged
virgin	92%	0%	0%
conditioned	0%	65%	0%
aged	0%	0%	83%

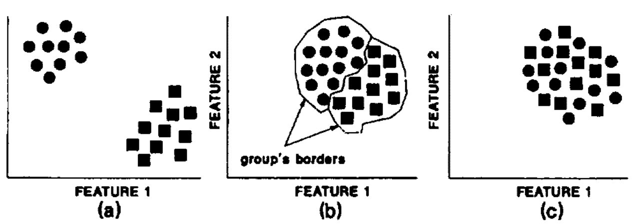

Figure 18.

(a) Two well-separated clusters. (b) One cluster consisting of two different groups. (c) One cluster consisting of two similar groups. Each dot represents one fingerprint.

It should be noted that clustering algorithms usually recognize clusters which are well separated, see Figure 18(a). Sometimes, however, different groups do not form separated clusters, although the groups occupy different positions in feature space. This situation is shown in Figure 18(b). In such cases, there is *no clustering* method which could distinguish between situations shown in Figures 18(b) and (c). In this case the use of mapping techniques would be more appropriate to discover the data structure. The *classification* of a fingerprint of unknown origin to two groups shown in Figures 18(a) and (b), can still be successful. The groups must then be known a priori, so that borders between the groups are known and satisfactory classification can take place. In the case shown in Figure 18(c) no difference can be found between the fingerprints of two groups.

Discovering structures in the data is a difficult task and there is no best method to perform this task [39, 144, 154, 187]. Each method examines the data in its own way, and it is best according to a criterion which the method optimizes [144, 154]. Scatter plots of fingerprints in two-dimensional space obtained by mapping techniques give the first impression of the structures in the data. Here, principal component analysis and discriminant analysis can each serve as a starting point. Other methods, such as multidimensional scaling and Sammon's mapping [197] are also recommended. Of the cluster analysis methods the group average method is preferred. It produces a tree structure which allows a detailed examination of the data.

There are, however, many problems related to discovering structures in data, such as the determination of the correct number of clusters in data, validation of individual clusters, validation of tree structures and so on. Numerous examples on the use of the methods can be found [144, 154, 185, 187, 221-224].

4. CLASSIFICATION OF FINGERPRINTS

The aim of classification is to assign a label to a PD pattern of unknown origin from previously collected patterns with known labels (treeing discharges, corona, *etc.*). A number of approaches and classification methods have been used in the past for PD recognition: expert systems [11, 12, 17], hidden Markov models [24], fuzzy logic [29, 225], neural networks [20-23, 26-29, 31-34, 36, 37] and conventional classification methods [15, 18, 25]. Each method has advantages and disadvantages in its use. To ensure that a classification method is suitable for a particular recognition task, it is especially important that a geometrical interpretation of the classification is understood. The border formation between fingerprints of different PD sources is explained with reference to examples of classification with neural networks (adaptive systems) and conventional classifiers (nonadaptive systems).

4.1. NEURAL NETWORKS

Neural networks (NN) have been successfully applied to a number of pattern classification problems [165, 166].

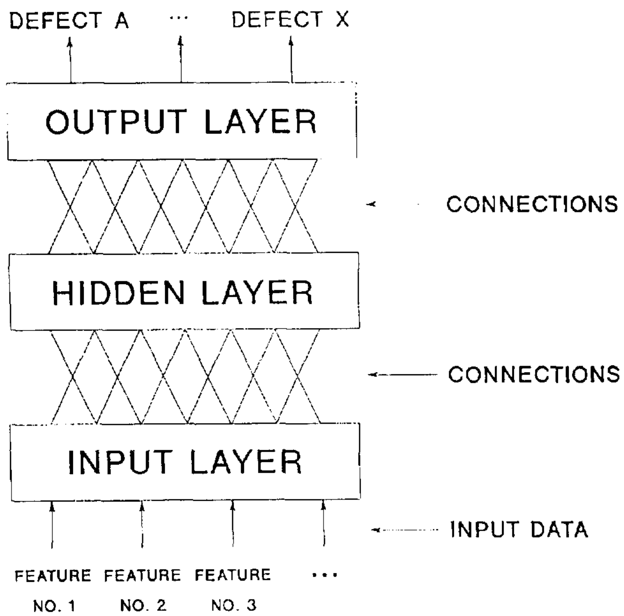

Figure 19.
General structure of a neural network.

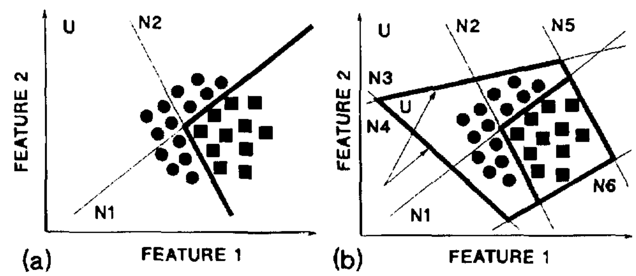

Figure 20.
Classification with the back-propagation neural network with: (a) two neurons in the hidden layer, (b) six neurons in the hidden layer. Hyperplanes (lines in 2-d space) are generated by the neurons N1 to N6 in the hidden layer and are shown by thin lines. Borders between the groups are determined by weight connections between the hidden and the output layer, and they are shown here by thick lines. The arrows in (b) show borders which were generated far away from the data. U: Fingerprint of unknown origin, ■: fingerprint of defect A, ●: fingerprint of defect B.

They have also been used in PD recognition, although with mixed results [20-23, 26-29, 31-34, 36, 37].

NN provides brain-like capabilities for solving problems: they learn by example [156, 157, 161, 165-173]. NN belong to nonparametric methods. This means that it is usually not necessary to make any assumptions about data structure. In statistics, various preliminary conditions, *e.g.* data from normal populations, must be fulfilled in order to carry out the analysis. The structure of NN is based on a mutually connected three-layer system, see Figure 19: input layer, hidden layer(s) and output layer.

The input layer may have several input neurons or processing elements, and is driven by values of features extracted from PD patterns. For example, the input data can be values of statistical parameters describing the shapes of $H_{qmax}(\varphi)$, $H_{qn}(\varphi)$, *etc.* distributions [20, 21, 23].

The hidden layer (or layers) characterizes the typical structure of a NN and is different for the diverse networks. The main purpose of the hidden layers is to extract classification information from the presented data.

The output layer is defined according to user expectations. It can be represented by one or more neurons whose outputs indicate the final classification of a PD pattern of unknown origin to the known patterns.

To assess the classification potential of an NN, it is important to understand how they classify fingerprints. Because there are many types of NN which classify fingerprints in a different way [226-242], only some of them are explained.

The back-propagation network (with one hidden layer and a sigmoid transfer function) separates data by hyperplanes (lines in 2-d space, planes in 3-d space, *etc.*). Detailed mathematical analysis can be found [227, 230, 231, 233, 236, 243]. The hyperplanes are generated by neurons in the hidden layer (one hyperplane per neuron). Weight connections between the input layer and the hidden layer determine the slope and shift of the hyperplanes. Weight connections between the hidden layer and the output layer serve as logical functions which decide on which side of a hyperplane a testing fingerprint is. This is shown in Figure 20(a). Fingerprints of two defects can be separated in this case by the network with two neurons in the hidden layer. A testing fingerprint is then classified according to its position relative to the hyperplanes. Such a classification procedure can, however, cause problems. It can be seen that a fingerprint of unknown origin U is in the present situation classified to be defect B, yet it apparently does not belong to the defect B. It follows that more neurons are required in the hidden layer to separate fingerprints of both defects from the surrounding space, see Figure 20(b). In this case, six neurons are used for such a separation of fingerprints. However, in more than two dimensions, the structure of data is unknown and it is difficult to estimate the number of neurons in the hidden layer. Furthermore, even if a sufficient number of neurons is supplied, the hyperplanes can still be generated far away from natural borders between the groups in the data so that a misclassification of a fingerprint of unknown origin can still occur, see Figure 20(b). Note that it does not really matter what kind of features are used as the input for NN (statistical parameters, fractal features, *etc.*) because the classification is determined by the classification principle of the network.

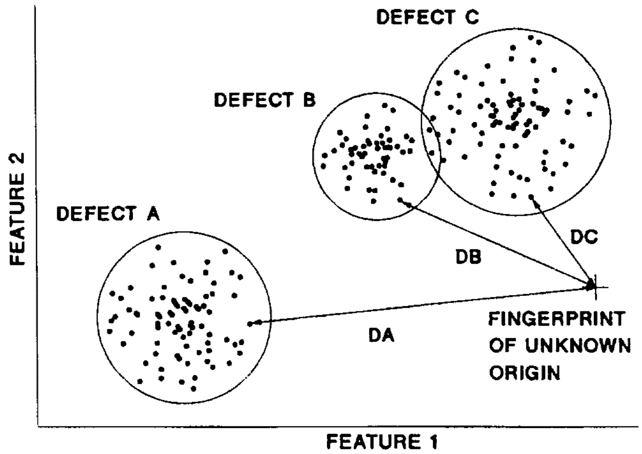

Figure 21.

The principle of the smallest distance classifier. One dot represents one fingerprint. A fingerprint of unknown origin is assigned to a defect with the smallest distance between the fingerprint and a defect.

The Kohonen self-organizing map, or learning vector quantization network are the smallest distance classifiers [232], *i.e.*, the networks classify a fingerprint of unknown origin to a defect with the smallest distance between the fingerprint and the defect, see Figure 21. This can cause problems in classification, because a fingerprint of unknown origin which is far away from reference data can still be classified as one of the known defects, as is shown in Figure 21. The use of these NN can be justified only if it is certain that a fingerprint of unknown origin belongs to one of the known categories.

An example of a fingerprint classification of actual PD data by the three types of NN described above is shown in Table 3. The input data were 15 values of statistical parameters as described in [26]. It can be seen that NN indeed classified correctly the fingerprints of defects they have been trained to recognize, in this case corona discharges in SF_6 and cavity discharges in a GIS spacer. However, a fingerprint of free conducting particle in GIS (with values of statistical parameters completely different from those of corona discharges in SF_6 and cavity discharges in the GIS spacer) was classified as cavity discharges in the spacer. The results confirm the classification principle of the back-propagation network, Kohonen self-organizing map and learning vector quantization network.

Some NN, such as the radial basis function network [229], fuzzy adaptive resonance theory (ART) network [237], restricted Coulomb energy network [226] can overcome the problems discussed above. For example, the

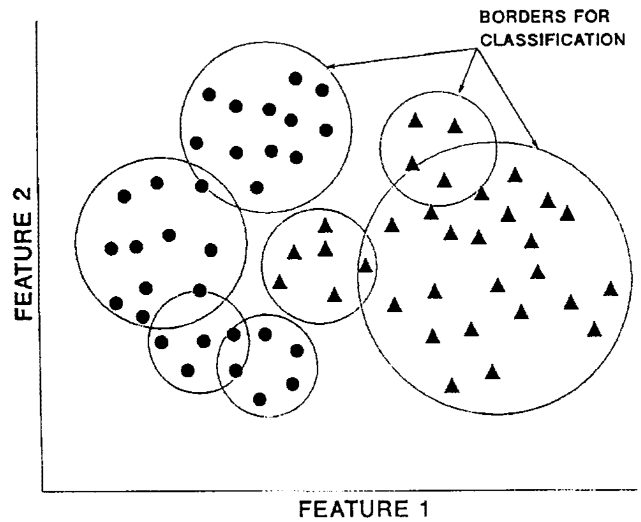

Figure 22.

Border formation between fingerprints of different PD sources in feature space by radial basis function neural network. ●: fingerprint of defect A, ▲: fingerprint of defect B.

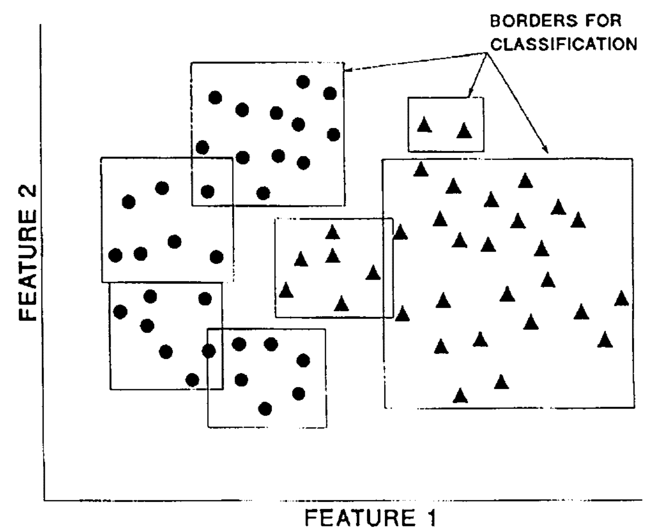

Figure 23.

Border formation between fingerprints of different PD sources in feature space by fuzzy Artmap neural network. ●: fingerprint of defect A, ▲: fingerprint of defect B.

radial basis function network encloses data in feature space by hyperspheres (circles in 2-d space, spheres in 3-d space, *etc.*) as shown in Figure 22. One hypersphere is generated by one neuron in the hidden layer. If a fingerprint of unknown origin falls outside a hypersphere, it is not assigned to any of the known defects. The fuzzy ART network encloses data in feature space by hypercubes (squares in 2-d space, cubes in 3-d space, *etc.*) as shown in Figure 23. Again if a fingerprint of unknown origin falls outside a hypercube, it is not assigned to any

Table 3.

Example of classifications with NN. Complete resemblance of a PD pattern of unknown origin to PD patterns of known defects is indicated by the value of 1, the complete lack of resemblance is indicated by the value of 0.

Fingerprint to be classified	Back propagation NN		Learning vector quantization NN		Kohonen self-organizing map	
Corona in SF_6	1	0	1	0	0.9	0
Cavity in GIS Spacer	0	1	0	1	0	0.9
Cond. part. in SF_6	0	0.9	0	1	0	0.9

of the known defects. It can be seen that these NN can provide crucial 'I do not know' answers and it would be interesting to apply the networks to PD recognition in the future.

There are, however, some other problems which complicate the use of NN in real applications. For example, the values of learning coefficients of NN have to be determined (however, only rough rules are available), classification can depend on a value of convergence criteria, learning times can be long, small sample size problems can occur [26, 184, 243]. Furthermore, many types of neural networks lack the modularity principle [239]. This principle says that one neural network trained to recognize, *e.g.* defects A and B, and another network trained to recognize defects C and D can be combined to a single network able to recognize defects A, B, C and D without additional learning. Usually, when there is a requirement of adding new defects for recognition, *e.g.* defects C and D, to a previously trained network (the one which recognizes defects A and B), the network has to be completely retrained. It should be kept in mind that the NN learning process can consume a lot of time, $\sim$ 3 h in these applications [20, 21, 26], although months of learning time have been reported in handwritten character recognition.

The progress in the NN field should be monitored carefully. When such types of NN have finally been designed that can overcome the problems mentioned here, their use for PD recognition will be justified.

4.2. CONVENTIONAL CLASSIFIERS

There are a number of conventional classifiers available for classification purposes: Bayes classifiers [39], Parzen classifiers [39, 244, 245], nearest neighbor classifiers [246], discriminant function classifiers [39, 154, 155], the centour score [160] and so on. In these methods, parameters such as the mean values, standard deviations of the features for different problems are first calculated. Then the distance between a fingerprint of unknown origin and, for instance, the mean value is calculated, *e.g.* by using the Pythagoras theorem. By statistical testing, such as χ^2 statistics, it can then be determined whether

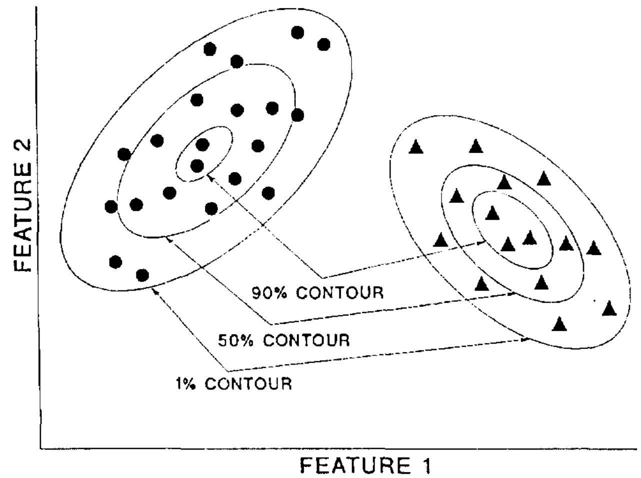

Figure 24.

Border formation with the centour score method. ●: fingerprint of defect A, ▲: fingerprint of defect B.

the distance between the fingerprint of unknown origin and a particular problem in a data base is small enough to assign the fingerprint to the problem. For example, the centour score method creates percentile contours in the form of hyperellipsoids (ellipses in 2-d space) around the mean values of features for a particular problem, see Figure 24. The size and the shape of hyperellipsoids is determined by the standard deviation of each feature and mutual correlation of the features [160]. If a fingerprint to be classified falls outside, *e.g.* the 1% contour, then the probability that the fingerprint belongs to this problem is $\leqslant$ 1%. It can then be concluded that the fingerprint does not belong to a given problem. If the values of features of a fingerprint to be classified are equal to the mean values of a particular defect, then the fingerprint belongs with 100% probability to the problem. Examples of classification with the centour score method are shown in Table 2. It can be seen that the classifications were correct in all cases. The centour method has been applied successfully to the recognition of artificially created defects in insulation and actual HV components [25, 48]. The use of the centour score is, however, restricted to normally distributed data of a particular defect. By the careful design

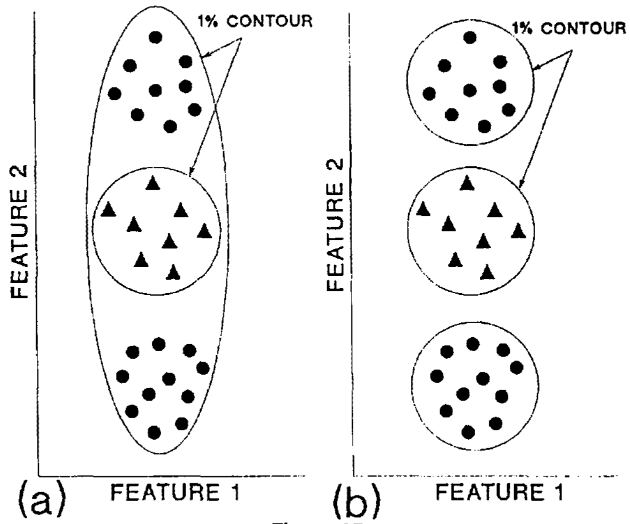

Figure 25.

(a) Border formation with the centour score method in the case of the supposition of normally distributed data of defect A. (b) Border formation with the centour score method when fingerprints of defect A were split into two normally distributed data. ●: fingerprint of defect A, ▲: fingerprint of defect B.

of a data base for discharge recognition, *e.g.* by splitting a non-normal distribution into several normal ones, this condition can be fulfilled so that there is no need to be afraid of misclassification. This situation is shown in Figure 25. If the fingerprints of defect A are considered as normally distributed data (which is obviously not true in this case) then the centour score (suitable for classification of normally distributed data) would wrongly estimate borders for the defect A, see Figure 25(a). When the fingerprints of defect A are split into two subgroups, then successful determination of borders (and thus classification) takes place with the centour score method. It can be seen from this example that the knowledge of data structure and the geometrical principle of border formation by a classification method are crucial points (among many others) in successful classification.

It should be noted that in pattern recognition the commonly used Bayes classifiers [39] assign a fingerprint of unknown origin to one of known categories. This is a serious disadvantage of the approach because the classifiers can not provide crucial 'I do not know' answers. The results obtained by various Bayes classifiers [218] resembled those shown in Table 3: a fingerprint which does not belong to any of the known problems was classified as one of the known problems.

There are of course many other, even more complicated, methods, such as the use of Parzen windows [39, 244, 245], potential functions [247], fuzzy classifiers [164], and abductive modeling methods [248-251]. The use of abductive modeling is particularly interesting. This technique attempts to find the best possible hypothesis to explain data. It has gained in popularity in the field of artificial intelligence in recent years, and it would be interesting to apply the method to PD recognition as well, especially for such complicated task as the recognition of multiple PD sources. This task has already been attempted (in a visual [43, 252-255] and an automated way [37]) with positive preliminary results, but PD patterns were classified with the back-propagation NN [37] which suffers from disadvantages as has been discussed above [26, 236, 243].

Because there are so many classification methods, it is important to select the correct one. When several competing classifiers are available, such as the centour score method, the back-propagation NN then the performance of each classifier should be assessed [39, 154, 256-258]. Some of the methods can easily be rejected by simple reasoning on their classification principle. For example, if it is required that a classification method must provide 'I do not know' answer as discussed above, then the minimum distance classifiers will hardly be a good choice. To obtain the statistical evaluation of the performance of a classifier, methods such as 'leave one out' can be used [259]. The procedure is as follows: the i-th fingerprint is deleted from a data set consisting of n fingerprints and the parameters such as the mean values and standard deviations, for a particular defect are calculated from remaining $(n-1)$ fingerprints. The deleted fingerprint is then classified to the collection of all defects and the response of a classification is noted, *e.g.* correct, incorrect, 'I do not know'. The whole procedure is then repeated for all fingerprints of all defects. By counting correct, incorrect, and 'I do not know' answers the performance of various classifiers can be estimated from the error count. The method has been used for the evaluation of the performance of the centour score method in PD recognition, [67]. The method is especially suitable for a data base with a small number of fingerprints. When the number of fingerprints in the data base increases the method can be time consuming. The method is known to produce an unbiased estimate of the error rate, although the estimate has a large variance. Some other methods for the estimation of the performance of classifiers and alternatives to this approach can be found in [154, 163, 258, 260-263].

5. CONCLUSIONS

IN this work an overview of automated PD recognition is given. It can be concluded that there are many ways to achieve this goal.

The first important step is to select a type of PD pattern that has good discriminating power. Especially $H_n(\varphi, q)$ PD distribution and its derivates such as $H_{qmax}(\varphi)$, $H_{qn}(\varphi)$, *etc.* distributions have been extensively used for recognition. The shape of individual PD pulses and various frequency spectra provide another way to recognize partial discharges.

To reduce the dimension of original PD data, $H_n(\varphi, q)$ distribution, 'features' or 'properties' of the data should be extracted from the data. There is no unique way to do this. Statistical parameters (skewness, kurtosis) and fractal features (fractal dimension, lacunarity) are just few examples of such features. The trade-off between the number of features, time for the calculation of the features, discriminating power of the features and the final speed of classification should be considered when designing the features.

To create a data base for reliable PD recognition, various aspects such as the effects of test voltage level, aging, availability of starting electrons, must be taken into account. A number of mathematical methods are available to organize the data base. Mapping techniques and cluster analysis methods can be used for this purpose but it should be realized that there is no 'best' method.

Many classification methods can be used for a final classification of a PD pattern of unknown origin to the known patterns. Satisfactory results were reported especially with the centour score method. NN produced mixed results in PD recognition. However, the progress in the NN field should be monitored. New types of NN might classify PD patterns more reliably. It would also be useful to apply methods such as abductive modeling for PD recognition in the future.

Results presented to date now dealt mostly with recognition of single PD sources. Future automated recognition systems should also be able to recognize multiple PD sources and to pinpoint the most dangerous one. Possibilities for monitoring the aging of insulation by means of PD recognition should also be further investigated.

ACKNOWLEDGMENT

The author wish to express his gratitude to the Department of Electrical Research and Information Technology (EOI) of N. V. Kema, Arnhem, the Netherlands for funding the research work.

REFERENCES

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

[2] S. Whitehead, *Dielectric Breakdown of Solids*, Clarendon Press, 1953.

[3] F. H. Kreuger, *Discharge Detection in High Voltage Equipment*, Temple Press, American Elsevier, 1964; Butterworths, 1989.

[4] R. Bartnikas, E. J. McMahon, Eds., *Engineering Dielectrics*, Vol. I, "Corona Measurement and Interpretation", ASTM, 1979.

[5] R. Bartnikas, E. J. McMahon, Eds., *Engineering Dielectrics*, Vol. IIa, "Electrical Properties of Solid Insulating Materials: Molecular Structure and Electrical Behavior", ASTM, 1983.

[6] L. A. Dissado, J. C. Fothergill, *Electrical Degradation and Breakdown in Polymers*, Peregrinus, 1992.

[7] IEC Publication 270, *Partial Discharge Measurements*, 1981.

[8] Working Group 21.03 Cigre, "Recognition of Discharges", Electra, Vol. 11, pp. 61–98, 1969.

[9] B. H. Ward, "Digital Techniques for Partial Discharge Measurements", IEEE Trans. on Power Delivery, Vol. 7, pp. 469–479, 1992.

[10] T. Okamoto and T. Tanaka, "Novel Partial Discharge Measurement Computer-aided Measurement Systems", IEEE Trans. on Electrical Insulation, Vol. 21, pp. 1015–1019, 1986.

[11] J. D. Gassaway, P. B. Jacob, C. A. Vassiliadis and P. H. Reynolds, "Computer-aided Partial Discharge Measurement and Recognition", 5th Int. Symp. on HV Engineering, Braunschweig, Germany, Paper 41.03, 1987.

[12] R. E. Wootton, "Computer Assistance for the Performance and Interpretation of High Voltage ac Discharge Tests", 5th Int. Symp. on HV Engineering, Braunschweig, Germany, Paper 41.12, 1987.

[13] E. Gulski, P. Morshuis and F. H. Kreuger, "Automized Recognition of Partial Discharges in Cavities", Japanese Journal of Applied Physics, Vol. 29, pp. 1329–1335, 1990.

[14] E. Gulski and F. H. Kreuger, "Computer-aided Analysis of Discharge Patterns", Journal of Physics D: Applied Physics, Vol. 23, pp. 1569–1575, 1990.

[15] E. Gulski, *Computer-aided Recognition of Partial Discharges Using Statistical Tools*, Delft University Press, 1991.

[16] E. Gulski and F. H. Kreuger, "Computer-aided Recognition of Discharge Patterns", IEEE Trans. on Electrical Insulation, Vol. 27, pp. 82–92, 1992.

[17] L. E. Lundgaard, G. Tangen, B. Skyberg and K. Faugstad, "Acoustic Diagnosis of Gis; Field Experience and Development of Expert System", IEEE Trans. on Power Delivery, Vol. 7, pp. 287–294, 1992.

[18] H.-G. Kranz and R. Krump, "Partial Discharge Diagnosis Using Statistical Optimization on a PC-

based System", IEEE Trans. on Electrical Insulation, Vol. 27, pp. 93–98, 1992.

[19] F. Komori, N. Nishiguchi, M. Hikita and T. Mizutani, "Construction of Prototype On-line System for Insulation Degradation Diagnosis of Polymeric Insulation Material Suffering Partial Discharges", 4th Int. Conf. on Conduction and Breakdown in Solid Dielectrics, Sestri Levante, Italy, pp. 155–159, 1992.

[20] H. Suzuki and T. Endoh, "Pattern Recognition of Partial Discharge in XLPE Cables Using a Neural Network", IEEE Trans. on Electrical Insulation, Vol. 27, pp. 543–549, 1992.

[21] N. Hozumi, T. Okamoto and T. Imajo, "Discrimination of Partial Discharge Patterns Using a Neural Network", IEEE Trans. on Electrical Insulation, Vol. 27, pp. 550–556, 1992.

[22] B. T. Phung, T. R. Blackburn and R. E. James, "The use of Artificial Neural Networks in Discriminating Partial Discharge Patterns", 6th Int. Conf. on Dielectric Materials, Measurements and Applications, Manchester, United Kingdom, pp. 25–28, 1992.

[23] A. Schnettler and V. Tryba, "Artificial Self-organizing Neural Network for Partial Discharge Recognition", Archiv für Elektrotechnik, Vol. 76, pp. 1–6, 1993.

[24] L. Satish and B. I. Gururaj, "Use of Hidden Markov Models for Partial Discharge Pattern Classification", IEEE Trans. on Electrical Insulation, Vol. 28, pp. 172–182, 1993.

[25] F. H. Kreuger, E. Gulski and A. Krivda, "Classification of Partial Discharges", IEEE Trans. on Electrical Insulation, Vol. 28, pp. 917–931, 1993.

[26] E. Gulski and A. Krivda, "Neural Networks as a Tool for Recognition of Partial Discharges", IEEE Trans. on Electrical Insulation, Vol. 28, pp. 984–1001, 1993. *Errata* in IEEE Trans. on Dielectrics and Electrical Insulation, Vol. 1, p. 351, 1994.

[27] H.-G. Kranz, "Diagnosis of Partial Discharge Signals Using Neural Networks and Minimum Distance Classifiers", IEEE Trans. on Electrical Insulation, Vol. 28, pp. 1016–1024, 1993.

[28] A. A. Mazroua, M. M. A. Salama and R. Bartnikas, "PD Pattern Recognition with Neural Networks Using the Multilayer Perceptron Technique", IEEE Trans. on Electrical Insulation, Vol. 28, pp. 1082–1089, 1993.

[29] A. Großand H.-G. Kranz, "Possibilities of Diagnostic Methods with Different Neural Networks and Fuzzy Sets for PD Measurements", 8th Int. Symp. on HV Engineering, Yokohama, Japan, Paper 62.12, 1993.

[30] F. H. Kreuger, P. H. F. Morshuis and E. Gulski, "Evaluation of Discharge Damage by Fast Transient Detection and Statistical Analysis", CIGRE, Paris, France, Paper 15–106, 1994.

[31] L. Satish and W. S. Zaengl, "Artificial Neural Networks for Recognition of 3-D Partial Discharge Patterns", IEEE Trans. on Dielectrics and Electrical Insulation, Vol. 1, pp. 265–275, 1994.

[32] A. A. Mazroua, R. Bartnikas and M. M. A. Salama, "Discrimination between Pd Pulse Shapes using Different Neural Network Paradigms", IEEE Trans. on Dielectrics and Electrical Insulation, Vol. 1, pp. 1119–1131, 1994.

[33] M. Oyama, E. Hanai, H. Aoyagi, H. Murase, I. Ohshima and S. Menju, "Development of Detection and Diagnostic Techniques for Partial Discharges in Gis", IEEE Trans. on Power Delivery, Vol. 9, pp. 811–818, 1994.

[34] A. A. Mazroua, R. Bartnikas and M. M. A. Salama, "Neural Network System Using the Multi-layer Perceptron Technique for the Recognition of PD Pulse Shapes due to Cavities and Electrical Trees", IEEE Trans. on Power Delivery, Vol. 10, pp. 92–96, 1995.

[35] L. Satish and W. S. Zaengl, "Can Fractal Features be used for Recognizing 3-D Partial Discharge Patterns?", IEEE Trans. on Dielectrics and Electrical Insulation, Vol. 2, pp. 352–359, 1995.

[36] H. Borsi, E. Gockenbach and D. Wenzel, "Separation of Partial Discharges from Pulse-shaped Noise Signals with the Help of Neural Networks", IEE Proc. on Science, Measurements and Technology, Vol. 142, pp. 69–74, 1995.

[37] T. Okamoto and T. Tanaka, "Partial Discharge Pattern Recognition for Three Kinds of Model Electrodes with a Neural Network", IEE Proc. on Science, Measurements and Technology, Vol. 142, pp. 75–84, 1995.

[38] T. Hücker and H.-G. Kranz, "New Approach in Partial Discharge Diagnosis and Pattern Recognition", IEE Proc. on Science, Measurements and Technology, Vol. 142, pp. 89- 94, 1995.

[39] R. O. Duda, P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, 1973.

[40] T. Okamoto, "Comprehensive Partial Discharge Characteristics Sensitive to Electrode Configuration", Conf. on Electrical Insulation and Dielectric Phenomena, pp. 126–131, 1987.

[41] M. Hikita, K. Yamada, A. Nakamura, T. Mizutani, A. Oohasi and M. Ieda, "Measurements of Partial Discharges by Computer and Analysis of Partial

Discharge Distribution by the Monte Carlo Method", IEEE Trans. on Electrical Insulation, Vol. 25, pp. 453–468, 1990.

[42] J. Fuhr, M. Hässig, B. Fruth and T. Kaiser, "PD-fingerprints of Some High Voltage Apparatus", IEEE Int. Symp. on Electrical Insulation, Toronto, Canada, pp. 129–132, 1990.

[43] B. Fruth and L. Niemeyer, "The Importance of Statistical Characteristics of Partial Discharge Data", IEEE Trans. on Electrical Insulation, Vol. 27, pp. 60–69, 1992.

[44] R. Bartnikas, "Note on Multichannel Corona Pulse-height Analysis", IEEE Trans. on Electrical Insulation, Vol. 8, pp. 2–5, 1973.

[45] A. Kelen, "The Functional Testing of High Voltage Generator Stator Insulation", CIGRE, Paris, France, Paper 15–03, 1976.

[46] T. Tanaka and T. Okamoto, "A Minicomputer-based Partial Discharge Measurement System", IEEE Int. Symp. on Electrical Insulation, Philadelphia, USA, pp. 86–89, 1978.

[47] R. Schifani, "A Novel Histogram for Partial Discharge Signals in High Voltage Insulating Systems", IEEE Trans. on Electrical Insulation, Vol. 21, pp. 89–99, 1986.

[48] E. Gulski, "Discharge Pattern Recognition in High Voltage Equipment", IEE Proc. on Science, Measurements and Technology, Vol. 142, pp. 51–61, 1995.

[49] G. C. Stone, H. G. Sedding, N. Fujimoto and J. M. Braun, "Practical Implementation of Ultrawideband Partial Discharge Detectors", IEEE Trans. on Electrical Insulation, Vol. 27, pp. 70–81, 1992.

[50] R. Bartnikas, "Corona Pulse Probability Density Function Measurements on Primary Distribution Power Cables", IEEE Trans. on Power Apparatus and Systems, Vol. 94, pp. 716–723, 1975.

[51] M. Kurtz and J. F. Lyles, "Generator Insulation Diagnostic Testing", IEEE Trans. on Power Apparatus and Systems, Vol. 98, pp. 1596–1603, 1979.

[52] D. M. Ryder, J. W. Wood and W. K. Hogg, "Changes in Discharge Amplitude Distribution with Dielectric Damage", 3rd Int. Conf. on Dielectric Materials, Measurements and Applications, Birmingham, United Kingdom, pp. 101–104, 1979.

[53] T. Okamoto and T. Tanaka, "Change of Pulse-mean $\varphi - q$ Distribution Patterns of Partial Discharges due to Treeing Propagation", Electrical Engineering in Japan, Vol. 102, pp. 18–25, 1982.

[54] W. K. Hogg, R. Miller, G. Rabach and D. M. Ryder, "The Relationship of Partial Discharge Amplitude Distributions with Dielectric Damage at Different Levels of Voltage and Frequency", IEEE Int. Symp. on Electrical Insulation, Philadelphia, USA, pp. 79–85, 1982.

[55] M. Henriksen, "Pulse Analysis on Partial Discharges in Gas Filled Cavities in Epoxy Resins", IEEE Int. Symp. on Electrical Insulation, Philadelphia, USA, pp. 86–90, 1982.

[56] T. Okamoto and T. Tanaka, "Prediction of Treeing Breakdown from Pulse Height of Partial Discharge on Voltage-phase Angle", Japanese Journal of Applied Physics, Vol. 24, pp. 156–160, 1985.

[57] G. Rabach, A. Contin and F. Tosato, "Study of Electrical Aging in Stator Bar Insulation by Partial Discharge Amplitude Distributions Analysis", 5th Int. Symp. on HV Engineering, Braunschweig, Germany, Paper 42.06, 1987.

[58] B. Fruth and J. Fuhr, "Partial Discharge Pattern Recognition - A Tool for Diagnosis and Monitoring of Aging", CIGRE, Paris, France, Paper 15/33–12, 1990.

[59] H. N. Geetha, M. B. Srinivas and T. S. Ramu, "Pulse-count Distribution as a Possible Diagnostic Tool for Assessing the Level of Degradation of Rotating Machine Insulation", IEEE Trans. on Electrical Insulation, Vol. 25, pp. 747–756, 1990. *Discussion* in IEEE Trans. on Electrical Insulation, Vol. 26, pp. 840–842, 1991.

[60] J. T. Holbøll and M. Henriksen, "Partial Discharge Patterns Related to Surface Deterioration in Voids in Epoxy", IEEE Int. Symp. on Electrical Insulation, Toronto, Canada, pp. 115–119, 1990.

[61] M. Pompili, C. Mazzetti and E. O. Forster, "Partial Discharge Distributions in Liquid Dielectrics", IEEE Trans. on Electrical Insulation, Vol. 27, pp. 99–105, 1992.

[62] Y. J. Kim and J. K. Nelson, "Assessment of Deterioration in Epoxy/mica Machine Insulation", IEEE Trans. on Electrical Insulation, Vol. 27, pp. 1026–1039, 1992.

[63] J. T. Holbøll and M. Henriksen, "Partial Discharge Patterns and Surface Deterioration in Voids in Filled and Unfilled Epoxy", IEEE Int. Symp. on Electrical Insulation, Baltimore, USA, pp. 354–358, 1992.

[64] A. Contin and G. Rabach, "Partial Discharge Analysis of Rotating ac Machines", IEEE Trans. on Electrical Insulation, Vol. 28, pp. 1033–1042, 1993.

[65] M. Hikita, F. Komori, N. Nishiguchi and T. Mizutani, "Phase-resolved and Time-sequential Partial Discharge Pulse Measurements in a Metal-void-insulator System Using a Personal Computer", Journal of Physics D: Applied Physics, Vol. 27,

pp. 1220–1228, 1994.

[66] R. J. Van Brunt, "Physics and Chemistry of Partial Discharge and Corona. Recent Advances and Future Challenges", IEEE Trans. on Dielectrics and Electrical Insulation, Vol. 1, pp. 761–784, 1994.

[67] A. Krivda and E. Gulski, "Influence of Aging on Classification of Partial Discharges in Cavities", Japanese Journal of Applied Physics, Part I, Vol. 33, pp. 5942–5949, 1994.

[68] A. C. Gjærde, "Measurements of Void Gas Pressure During Combined Thermal and Partial Discharge Aging of Epoxy", IEE Proc. on Science, Measurements and Technology, Vol. 142, pp. 17–21, 1995.

[69] S. A. Boggs, G. L. Ford and R. C. Madge, "Coupling Devices for the Detection of Partial Discharges in Gas Insulated Switchgear", IEEE Trans. on Power Apparatus and Systems, Vol. 100, pp. 3969–3973, 1981.

[70] S. A. Boggs, "Electromagnetic Techniques for Fault and Partial Discharge Location in Gas Insulated Cables and Substations", IEEE Trans. on Power Apparatus and Systems, Vol. 101, pp. 1935–1941, 1982.

[71] S. R. Campbell, G. S. Stone and H. G. Sedding, "Application of Pulse Width Analysis to Partial Discharge Detection", IEEE Int. Symp. on Electrical Insulation, Baltimore, USA, pp. 345–348, 1992.

[72] J. C. Devins, "Current Pulse Shapes of Discharges in Air Gaps Limited by Series Dielectric", Conf. on Electrical Insulation and Dielectric Phenomena, pp. 6–11, 1958.

[73] P. H. F. Morshuis and F. H. Kreuger, "Transition from Streamer to Townsend Mechanisms in Dielectric Voids", Journal of Physics D: Applied Physics, Vol. 23, pp. 1562–1568, 1990.

[74] P. H. F. Morshuis and F. H. Kreuger, "The Spatial Distribution and Electrical Parameters of Partial Discharges in Polyethylene Insulation during Aging", 4th Int. Conf. on Conduction and Breakdown in Solid Dielectrics, Sestri Levante, Italy, pp. 209–214, 1992.

[75] C. Hudon, R. Bartnikas and M. R. Wertheimer, "Spark-to-glow Discharge Transition due to Increased Surface Conductivity on Epoxy Resin Specimens", IEEE Trans. on Electrical Insulation, Vol. 28, pp. 1–8, 1993.

[76] P. H. F. Morshuis, *Partial Discharge Mechanisms*, Delft University Press, 1993.

[77] H. Dejean, P. P. F. Dejean and M. Goldman, "On Site Very High Frequency PD Pulse Measurements on Full Size Rotating Machine Stator for Discharge Type Recognition", Int. Conf. on Partial Discharge, Canterbury, United Kingdom, pp. 66–67, 1993.

[78] Y. Mizuno, J. M. Braun, R. J. Densley, N. Fujimoto and H. G. Sedding, "Temporal Evolution of Partial Discharge in Voids and Electrical Trees", 8th Int. Symp. on HV Engineering, Yokohama, Japan, Paper 60.03, 1993.

[79] P. H. F. Morshuis, "Partial Discharge Mechanisms in Voids Related to Dielectric Degradation", IEE Proc. on Science, Measurements and Technology, Vol. 142, pp. 62- 68, 1995.

[80] T. Kalicki, J. M. Braun, J. Densley, N. Fujimoto and H. G. Sedding, "Partial Discharge Pulse Characteristics of Electrical Trees in XLPE Cable Insulation", 5th Int. Conf. on Conduction and Breakdown in Solid Dielectrics, Leicester, England, pp. 528–532, 1995.

[81] R. J. Van Brunt and D. Leep, "Characterization of Point-plane Corona Pulses in SF_6", Journal of Applied Physics, Vol. 52, pp. 6588–6600, 1981.

[82] H. F. A. Verhaart and P. C. T. van Der Laan, "Fast Current Measurements for Avalanche Studies", Journal of Applied Physics, Vol. 53, pp. 1430–1436, 1982.

[83] J. C. Devins, "The Physics of Partial Discharges in Solid Dielectrics", IEEE Trans. on Electrical Insulation, Vol. 19, pp. 475–495, 1984.

[84] H. F. A. Verhaart and P. C. T. van Der Laan, "The Influence of Water Vapor on Avalanches in Air", Journal of Applied Physics, Vol. 55, pp. 3286–3292, 1984.

[85] S. V. Kulkarni and R. S. Nema, "Broad Band Pulse Detection Studies of Corona and Breakdown in Air, N_2, O_2, CO_2, SF_6 and SF_6-N_2 Mixtures", in *Gaseous Dielectrics V*, eds. L. G. Christophorou, D. W. Bouldin, Pergamon Press, pp. 637–642, 1987.

[86] M. Černák and T. Hosokawa, "Similarities between the Initial Phase of a Transient Nonuniform Glow Discharge in Nitrogen and the Negative Corona Trichel Pulse Formation in an Electronegative Gas", Applied Physics Letters, Vol. 52, pp. 185–187, 1988.

[87] J. M. Wetzer and P. C. T. van Der Laan, "Prebreakdown Currents", IEEE Trans. on Electrical Insulation, Vol. 24, pp. 297–308, 1989.

[88] R. J. Van Brunt, "Stochastic Properties of Partial-discharge Phenomena", IEEE Trans. on Electrical Insulation, Vol. 26, pp. 902–948, 1991.

[89] M. Černák and T. Hosokawa, "Complex Form of Current Pulses in Negative Corona Discharges", Physical Review A, Vol. 43, pp. 1107–1109, 1991.

[90] S. A. Boggs and G. C. Stone, "Fundamental Limitations in the Measurement of Corona and Partial Discharge", IEEE Trans. on Electrical Insulation, Vol. 17, pp. 143–150, 1982.

[91] A. Wilson, R. J. Jackson and N. Wang, "Discharge Detection Techniques for Stator Windings", IEE Proc., Part B, Vol. 132, pp. 234–244, 1985.

[92] M. Henriksen, G. S. Stone and M. Kurtz, "Propagation of Partial Discharge and Noise Pulses in Turbine Generator", IEEE Trans. on Energy Conversion, Vol. 1, pp. 161–166, 1986.

[93] J. W. Wood, H. G. Sedding, W. K. Hogg, I. J. Kemp and H. Zhu, "Partial Discharges in HV Machines; Initial Considerations for a PD Specification", IEE Proc., Part A, Vol. 140, pp. 409–416, 1993.

[94] R. J. Van Brunt, "Importance of Unraveling Memory Propagation Effects in Interpreting Data on Partial Discharge Statistics", IEEE Trans. on Electrical Insulation, Vol. 28, pp. 905–916, 1993.

[95] J. C. Bapt, Bui Ai and C. Mayoux, "Corona Frequency Analysis in Artificial Cavities in Epoxy Resins", Conf. on Electrical Insulation and Dielectric Phenomena, pp. 282–288, 1973.

[96] M. Hoof and R. Patsch, "Pulse-sequence Analysis: A New Method for Investigating the Physics of PD-induced Aging", IEE Proc. on Science, Measurements and Technology, Vol. 142, pp. 95–101, 1995.

[97] U. Fromm, *PD and Breakdown Testing of High Voltage dc Equipment*, Delft University Press, 1995.

[98] U. Fromm, "Partial Discharges at dc Voltages", to appear in IEEE Trans. on Dielectrics and Electrical Insulation.

[99] B. F. Hampton and R. J. Meats, "Diagnostic Measurements at UHF in Gas Insulated Substations", Proc. IEE, Part C, Vol. 135, pp. 137–144, 1988.

[100] J. S. Pearson, B. F. Hampton and A. G. Sellars, "A Continuous Uhf Monitor for Gas-insulated Substations", IEEE Trans. on Electrical Insulation, Vol. 26, pp. 469–478, 1991.

[101] A. Bargigia, W. Koltunowicz and A. Pigini, "Detection of Partial Discharges in Gas Insulated Substations", IEEE Trans. on Power Delivery, Vol. 7, pp. 1239–1249, 1992.

[102] T. Ozaki, K.-I. Abe and T. Umemura, "Partial Discharge Detection using Ferrite Antenna", IEEE Int. Symp. on Electrical Insulation, Baltimore, USA, pp. 371–374, 1992.

[103] A. G. Sellars, O. Farish and M. M. Peterson, "UHF Detection of Leader Discharges in SF_6", IEEE Trans. on Dielectrics and Electrical Insulation, Vol. 2, pp. 143–154, 1995.

[104] R. Arora, P. K. Kalra and R. K. Srivastava, "An Investigation of Electromagnetic Interference Caused by Different Types of Coronas in Air", IEEE Int. Symp. on Electrical Insulation, Baltimore, USA, pp. 287–290, 1992.

[105] R. T. Harrold, F. T. Emery, F. J. Murphy and S. A. Drinkut, "Radio Frequency Sensing of Incipient Arcing Faults within Large Turbine Generators", IEEE Trans. on Power Apparatus and Systems, Vol. 98, pp. 1167–1173, 1979.

[106] F. T. Emery and R. T. Harrold, "On Line Incipient Arc Detection in Large Turbine Generators Stator Windings", IEEE Trans. on Power Apparatus and Systems, Vol. 99, pp. 2232–2240, 1980.

[107] F. T. Emery, B. N. Lenderking and R. D. Couch, "Turbine - Generator On-line Diagnostics Using RF Monitoring", IEEE Trans. on Power Apparatus and Systems, Vol. 100, pp. 4974–4982, 1981.

[108] J. E. Timperley, "Incipient Fault Identification Through Neutral RF Monitoring of Large Rotating Machines", IEEE Trans. on Power Apparatus and Systems, Vol. 102, pp. 693–698, 1983.

[109] P. Grünewald, J. Weidner and A. Wichmann, "Radio Frequency Monitoring of the Windings of Electrical Machines", Siemens Power Engineering and Automation, Vol. 7, No. 1, pp. 26–29, 1985.

[110] F. T. Emery and R. T. Harrold, "Radio Frequency Response of a Large Turbine-generator Stator Winding", IEEE Trans. on Energy Conversion, Vol. 1, pp. 172–179, 1986.

[111] J. E. Timperley and E. K. Chambers, "Locating Defects in Large Rotating Machines and Associated Electrical Systems Through EMI Diagnostics", CIGRE, Paris, France, Paper 11–311, 1992.

[112] A. G. Sellars, O. Farish and B. F. Hampton, "Characterising the Discharge Development due to Surface Contamination in GIS using the UHF Technique", IEE Proc. on Science, Measureemnt and Technology, Vol. 141, pp. 118–122, 1994.

[113] A. G. Sellars, O. Farish and B. F. Hampton, "Assessing the Risk of Failure due to Particle Contamination of GIS using the UHF Technique", IEEE Trans. on Dielectrics and Electrical Insulation, Vol. 1, pp. 323–331, 1994.

[114] R. T. Harrold, "Ultrasonic Spectrum Signatures of Under-oil Corona Sources", IEEE Trans. on Electrical Insulation, Vol. 10, pp. 109–112, 1975.

[115] R. T. Harrold, "Acoustical Technology Applications in Electrical Insulation and Dielectrics", IEEE Trans. on Electrical Insulation, Vol. 20, pp.

3–19, 1985.

[116] N. Okutsu, S. Matsuda, H. Mukae, Y. Takahashi and S. Tominaga, "Pattern Recognition of Vibrations in Metal Enclosures of Gas Insulated Equipment and Its Application", IEEE Trans. on Power Apparatus and Systems, Vol. 100, pp. 2733–2739, 1981.

[117] H. Ogi, H. Tanaka, Y. Akimoto, Y. Izui, H. Taoka and T. Sakaguchi, "Preventive Maintance System for Gas Insulated Switchgear using an Artificial Neural Network", 3rd Symp. on Expert Systems Application to Power Systems, Tokyo-Kobe, Japan, pp. 627–633, 1991.

[118] E. Howells and E. T. Norton, "Detection of Partial Discharges in Transformers Using Acoustic Emission Techniques", IEEE Trans. on Power Apparatus and Systems, Vol. 97, pp. 1538–1548, 1978.

[119] H. Fujita, T. Nakanishi and K. Yamaguchi, "Acoustic Emission Distributions and Types of Electrical Trees in Polyester Resin", IEEE Trans. on Electrical Insulation, Vol. 18, pp. 520–527, 1983.

[120] B. A. Fruth and D. W. Gross, "Partial Discharge Signal Generation Transmission and Acquisition", IEE Proc. on Science, Measurements and Technology, Vol. 142, pp. 22- 28, 1995.

[121] M. A. S. Leijon and A. E. Vlastõs, "Pattern Recognition of Free Metallic Motion Modes in GIS", IEEE Int. Symp. on Electrical Insulation, Boston, USA, pp. 120–123, 1988.

[122] B. S. Hansen and F. Levring, "Optical Investigations of the Spatial and Temporal Development of Partial Discharges in Polyethylene", IEEE Int. Symp. on Electrical Insulation, Philadelphia, USA, pp. 296–299, 1982.

[123] B. Gravendeel, F. J. De Hoog and M. A. M. Schoenemakers, "Fast Photon Counting in Negative Corona Discharges in the Trichel Regime", Journal of Physics D: Applied Physics, Vol. 21, pp. 744–755, 1988.

[124] F. Grum and L. F. Costa, "Spectral Emission of Corona Discharges", Applied Optics, Vol. 15, pp. 76–79, 1976.

[125] N. Yoshimura, M. Nishida and F. Noto, "Light Emission from Tracking Discharges on Organic Insulation", IEEE Trans. on Electrical Insulation, Vol. 19, pp. 149–155, 1984.

[126] R. Mangeret, J. Farenc, Bui Ai, P. Destruel, D. Puertolas and J. Casanovas, "Optical Detection of Partial Discharges using Fluorescent Fiber", IEEE Trans. on Electrical Insulation, Vol. 26, pp. 783–789, 1991.

[127] K. Fujii, M. Yamada, A. Tanaka and K. Kurosawa, "Emission Spectrum of Partial Discharge in SF_6 Gas", IEEE Int. Symp. on Electrical Insulation, Baltimore, USA, pp. 332–335, 1992.

[128] Y. Shibuya, S. Zoledziowski and J. H. Calderwood, "Light Emission and Deterioration in Epoxy Resin Subjected to Power Frequency Electric Fields", Proc. IEE, Vol. 125, pp. 352–354, 1978.

[129] N. Shimizu, H. Katsukawa, M. Miyauchi, M. Kosaki and K. Horii, "The Space Charge Behavior and Luminescence Phenomena in Polymers at 77 K", IEEE Trans. on Electrical Insulation, Vol. 14, pp. 256–262, 1979.

[130] K. Nakanishi and S. Hirabayashi, "Phenomena and Mechanisms of Tree Inception in Epoxy Resins", IEEE Trans. on Electrical Insulation, Vol. 14, pp. 306–314, 1979.

[131] C. Laurent, C. Mayoux and S. Noel, "Dielectric Breakdown of Polyethylene in Divergent Field: Role of Dissolved Gases and Electroluminescence", Jornal of Applied Physics, Vol. 54, pp. 1532–1539, 1983.

[132] Y. Kitamura and S. Hirabayashi, "Partial Discharge Deterioration of Epoxy Resin for Electronic Parts", Conf. on Electrical Insulation and Dielectric Phenomena, Buffalo, USA, pp. 485–490, 1985.

[133] S. S. Bamji, A. T. Bulinski and R. J. Densley, "The Role of Polymer Interface during Tree Initiation in LDPE", IEEE Trans. on Electrical Insulation, Vol. 21, pp. 639–644, 1986.

[134] F. H. Kreuger, P. H. F. Morshuis and W. A. Sonneveld, "Optical Detection of Surface Discharges", IEEE Trans. on Electrical Insulation, Vol. 23, pp. 447–449, 1988.

[135] S. S. Bamji, A. T. Bulinski and R. J. Densley, "Threshold Voltage of Luminescence and Electrical Tree Inception in Low-density Polyethylene", Journal of Applied Physics, Vol. 63, pp. 5841–5845, 1988.

[136] S. S. Bamji, A. T. Bulinski and R. J. Densley, "Degradation of Polymeric Insulation Due to Photoemission Caused by High Electric Fields", IEEE Trans. on Electrical Insulation, Vol. 24, pp. 91–98, 1989.

[137] J. V. Champion, S. J. Dodd and G. C. Stevens, "Quantitative Measurement of Light Emission during the Early Stages of Electrical Breakdown in Epoxy and Unsaturated Polyester Resins", Journal of Physics D: Applied Physics, Vol. 26, pp. 819–828, 1993.

[138] R. Bartnikas, "A Commentary on Partial Discharge Measurement and Detection", IEEE Trans. on Electrical Insulation, Vol. 22, pp. 629–653, 1987.

[139] R. Bartnikas, "Detection of Partial Discharges (Corona) in Electrical Apparatus", IEEE Trans. on Electrical Insulation, Vol. 25, pp. 111–124, 1990.

[140] W. S. Zaengl, K. Lehmann and M. Albiez, "Conventional Pd Measurement Techniques used for Complex High Voltage Apparatus", IEEE Trans. on Electrical Insulation, Vol. 27, pp. 15–27, 1992.

[141] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, 1972, 1990.

[142] C.-H. Chen, *Statistical Pattern Recognition*, Spartan Books, 1973.

[143] J. T. Tou, R. C. Gonzales, *Pattern Recognition Principles*, Addison-wesley, 1974.

[144] A. K. Jain, R. C. Dubes, *Algorithms for Clustering Data*, Prentice Hall, 1988.

[145] P. S. Mohahir, *Pattern Recognition Transforms*, Research Studies Press, 1992.

[146] P. Morshuis, "Ultra-wide Band PD Detection and the Assessment of Dielectric Degradation", to be published in IEEE Trans. on Dielectrics and Electrical Insulation.

[147] X. Wu, H. Xie and S. Wu, "Application of Wavelet Analysis for Detection of Defects in Insulation Systems", 5th Int. Conf. on Conduction and Breakdown in Solid Dielectrics, Leicester, England, pp. 229–233, 1995.

[148] D. Wenzel, H. Borsi and E. Gockenbach, "A New Method of Partial Discharge Localization in Transformers via the Karhunen-Loève Transform", 4th Int. Conf. on Properties and Applications of Dielectric Materials, Brisbane, Australia, pp. 622–625, 1994.

[149] P. H. A. Sneath, R. R. Sokal, *Numerical Taxonomy: The Principles and Practice of Numerical Classification*, Freeman and Company, 1973.

[150] P. M. Mather, *Computational Methods of Multivariate Analysis in Physical Geography*, Wiley, 1976.

[151] C. Chatfield, A. J. Collins, *Introduction to Multivariate Analysis*, Chapman-hall, 1980.

[152] K. V. Mardia, J. T. Kent, J. M. Bibby, *Multivariate Data Analysis*, Academic Press, 1980.

[153] B. S. Everitt, G. Dunn, *Advanced Methods of Data Exploration and Modeling*, Heinemann Educational Books, 1983.

[154] W. Krzanowski, *Principles of Multivariate Analysis*, A User's Perspective, Clarendon Press, 1988.

[155] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley, 1958.

[156] G. S. Sebestyen, *Decision-making Processes in Pattern Recognition*, MacMillan, 1962.

[157] N. J. Nilsson, *Learning Machines*, McGraw-hill, 1965.

[158] A. G. Arkadev, E. M. Braverman, *Teaching Computers to Recognize Patterns*, Academic Press, 1967.

[159] E. A. Patrick, *Fundamentals of Pattern Recognition*, Prentice-hall, 1972.

[160] R. H. Lindeman, P. F. Merenda, R. Z. Gold, *Introduction to Bivariate and Multivariate Analysis*, Scott, Foresman and Company, 1980.

[161] J. Sklansky, G. N. Wassel, *Pattern Classifiers and Trainable Machines*, Springer Verlag, 1981.

[162] D. J. Hand, *Kernel Discriminant Analysis*, Research Studies Press, 1982.

[163] P. A. Devijver, J. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice-hall, 1982.

[164] S. K. Pal, D. K. D. Majumder, *Fuzzy Mathematical Approach to Pattern Recognition*, Wiley, 1986.

[165] J. A. Anderson, E. Rosenfeld, *Neurocomputing 1*, MIT Press, 1988.

[166] J. A. Anderson, A. Pellionisz, E. Rosenfeld, *Neurocomputing 2*, MIT Press, 1990.

[167] R. C. Eberhart, R. W. Dobbins, *Neural Network PC Tools*, Academic Press, 1990.

[168] J. A. Freeman, D. M. Skapura, *Neural Networks*, Addison-Wesley, 1991.

[169] R. Hecht-Nielsen, *Neurocomputing*, Addison-Wesley, 1991.

[170] R. J. Mammone, Y. Y. Zeevi, *Neural Networks*, Academic Press, 1991.

[171] P. D. Wasserman, *Advanced Methods in Neural Computing*, Van Nostrand Reinhold, 1993.

[172] S. Y. Kung, *Digital Neural Networks*, Prentice Hall, 1993.

[173] S. I. Gallant, *Neural Network Learning and Expert Systems*, MIT Press, 1993.

[174] J. Fuhr, B. Fruth, L. Niemeyer, D. Königstein, M. Hässig and F. Gutfleisch, "Generic Procedure for Classification of Aged Insulating System", Proc. of the 3rd Int. Conf. on Properties and Applications of Dielectric Materials, Tokyo, Japan, pp. 35–38, 1991.

[175] T. R. Dyckman, S. Smidt, A. K. McAdams, *Management Decision Making under Uncertainity*, The MacMillan Company, 1969.

[176] D. J. White, *Fundamentals of Decision Theory*, American Elsevier, 1976.

[177] C.-l. Hwang, K. Yoon, *Multiple Attribute Decision Making*, Springer Verlag, 1981.

[178] A. N. Mucciardi and E. E. Gose, "A Comparision of Seven Techniques for Choosing Subsets of Pattern Recognition Properties", IEEE Trans. on Computers, Vol. 20, pp. 1023–1031, 1971.

[179] C.-y. Chang, "Dynamic Programming as Applied to Feature Subset Selection in a Pattern Recognition System", IEEE Trans. on Systems, Man, and Cybernetics, Vol. 3, pp. 166–171, 1973.

[180] P. M. Narendra and K. Fukunaga, "A Branch and Bound Algorithm for Feature Subset Selection", IEEE Trans. on Computers, Vol. 26, pp. 917–922, 1977.

[181] T. M. Cover and J. M. Van Campenhout, "On the Possible Orderings in the Measurement Selection Problem", IEEE Trans. on Systems, Man, and Cybernetics, Vol. 7, pp. 657- 661, 1977.

[182] S. J. Raudys and V. Pikelis, "On Dimensionality, Sample Size, Classification Error, and Complexity of Classification Algorithm in Pattern Recognition", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 2, pp. 242–252, 1980.

[183] S. J. Raudys and A. K. Jain, "Small Sample Size Effects in Statistical Pattern Recognition: Recommendations for Practitioners", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 13, pp. 252–264, 1991.

[184] S. J. Raudys and A. K. Jain, "Small Sample Size Problems in Designing Artificial Neural Networks", in *Artificial Neural Networks and Statistical Pattern Recognition, Old and New Connections*, eds. I. K. Sethi, A. K. Jain, pp. 33–50, Elsevier Science Publishers, 1991.

[185] M. R. Anderberg, *Cluster Analysis for Applications*, Academic Press, 1973.

[186] H. Späth, *Cluster Analysis Algorithms*, Ellis Horwood, 1980.

[187] B. Everitt, *Cluster Analysis*, SSRC-Halsted Press, 1980.

[188] A. D. Gordon, *Classification*, Chapman-Hall, 1981.

[189] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, 1981.

[190] H. C. Romesburg, *Cluster Analysis for Researchers*, Wadsworth, 1984.

[191] L. Kaufman, P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, 1990.

[192] M. Jambu, *Exploratory and Multivariate Data Analysis*, Academic Press, 1991.

[193] J. C. Bezdek, S. K. Pal, *Fuzzy Models for Pattern Recognition*, IEEE Press, 1992.

[194] R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems", Annals of Eugenics, Vol. 3, pp. 179–188, 1936.

[195] J. B. Kruskal, "Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis", Psychometrika, Vol. 29, pp. 1–27, 1964.

[196] J. B. Kruskal, "Nonmetric Multidimensional Scaling: A Numerical Method", Psychometrika, Vol. 29, pp. 115–129, 1964.

[197] J. W. Sammon, "A Nonlinear Mapping for Data Structure Analysis", IEEE Trans. on Computers, Vol. 18, pp. 401–409, 1969.

[198] K. Fukunaga and W. L. Koontz, "Application of the Karhunen-Loève Expansion to Feature Selection and Ordering", IEEE Trans. on Computers, Vol. 19, pp. 311–318, 1970.

[199] C. L. Chang and R. C. T. Lee, "A Heuristic Method for Nonlinear Mapping in Cluster Analysis", IEEE Trans. on Systems, Man and Cybernetics, Vol. 3, pp. 197–200, 1973.

[200] J. H. Friedman and J. W. Tukey, "A Projection Pursuit Algorithm for Exploratory Data Analysis", IEEE Trans. on Computers, Vol. 23, pp. 881–889, 1974.

[201] R. C. T. Lee, J. R. Slagle and H. Blum, "A Triangulation Method for the Sequential Mapping of Points from N-space to Two-space", IEEE Trans. on Computers, Vol. 26, pp. 288–292, 1977.

[202] H. Niemann and J. Weiss, "A Fast-converging Algorithm for Nonlinear Mapping of High-Dimensional Data to a Plane", IEEE Trans. on Computers, Vol. 28, pp. 142–147, 1979.

[203] G. Biswas, A. K. Jain and R. C. Dubes, "Evaluation of Projection Algorithms", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 3, pp. 701–708, 1981.

[204] D. F. Mix and R. A. Jones, "A Dimensionality Reduction Technique Based on a Least Squared Error Criterion", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 4, pp. 537–544, 1982.

[205] W. Siedlecki, K. Siedlecka and J. Sklansky, "Mapping Techniques for Exploratory Pattern Analysis", in *Pattern Recognition and Artificial Intelligence*, eds. E. S. Gelsema, L. N. Kanal, Elsevier, pp. 277–299, 1988.

[206] J. Mao and A. K. Jain, "Artificial Neural Networks for Feature Extraction and Multivariate Data Projection", IEEE Trans. on Neural Networks, Vol. 6, pp. 296–317, 1995.

[207] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations", in *Proceedings of the 5th Berkeley Symposium on*

Mathematical Statistics and Probability, eds. L. M. Le Cam, J. Neyman, Vol. 1, pp. 281–297, University of California Press, 1967.

[208] C. T. Zahn, "Graph-theoretical Methods for Detecting and Describing Gestalt Clusters", IEEE Trans. on Computers, Vol. 20, pp. 68–86, 1971.

[209] W. L. G. Koontz and K. Fukunaga, "A Nonparametric Valley-seeking Technique for Cluster Analysis", IEEE Trans. on Computers, Vol. 21, pp. 171–178, 1972.

[210] R. A. Jarvis and E. A. Patrick, "Clustering Using a Similarity Measure Based on Shared Near Neighbors", IEEE Trans. on Computers, Vol. 22, pp. 1025–1034, 1973.

[211] D. E. Gustafson and W. Kessel, "Fuzzy Clustering with a Fuzzy Covariance Matrix", Proceedings of the 1978 IEEE Conference on Decision and Control, San Diego, USA, pp. 761–766, 1979.

[212] E. Trauwaert, "L_1 in Fuzzy Clustering", in *Statistical Data Analysis Based on the L_1 Norm and Related Methods*, ed. Dodge, Y., pp. 417–426, Elsevier Science Publishers, 1987.

[213] J. C. Gower, "A Comparision of Some Methods of Cluster Analysis", Biometrics, Vol. 23, pp. 623–637, 1967.

[214] R. M. Cormack, "A Review of Classification", Journal of the Royal Statistical Society, Vol. A134, pp. 321–367, 1971.

[215] F. K. Kuiper and L. Fisher, "A Monte Carlo Comparision of Six Clustering Procedures", Biometrics, Vol. 31, pp. 777–783, 1975.

[216] J. A. Hartigan, "Statistical Theory in Clustering", Journal of Classification, Vol. 2, pp. 63–76, 1985.

[217] G. N. Lance and W. T. Williams, "A General Theory of Classificatory Sorting Strategies, 1. Hierarchical Systems", Computer Journal, Vol. 9, pp. 373–380, 1966.

[218] A. Krivda, *Recognition of Discharges. Discrimination and Classification*, Delft University Press, 1995.

[219] A. Krivda and E. Gulski, "Classification of Discharge Patterns during Aging of Insulation", 5th Int. Conf. on Conduction and Breakdown in Solid Dielectrics, Leicester, England, pp. 254–258, 1995.

[220] A. Krivda and E. Gulski, "Influence of Aging on Classification of Partial Discharges", 9th Int. Symp. on HV Engineering, Graz, Austria, Paper 5643, 1995.

[221] C. Edelbrock, "Mixture Model Tests of Hierarchical Clustering Algorithms: The Problem of Classifying Everybody", Multivariate Behavioral Research, Vol. 14, pp. 367–384, 1979.

[222] G. W. Milligan, "An Examination of the Effect of Six Types of Error Perturbation on Fifteen Clustering Algorithms", Psychometrika, Vol. 45, pp. 325–342, 1980.

[223] G. W. Milligan, S. C. Soon and L. M. Sokol, "The Effect of Cluster Size, Dimensionality and the Number of Clusters on Recovery of True Cluster Structure", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 5, pp. 40–47, 1983.

[224] G. W. Milligan and M. C. Cooper, "An Examination of Procedures for Determining the Number of Clusters in a Data Set", Psychometrika, Vol. 50, pp. 159–179, 1985.

[225] D. Wenzel, H. Borsi and E. Gockenbach, "Partial Discharge Recognition and Localization in Transformers via Fuzzy Logic", IEEE Int. Symp. on Electrical Insulation, Pittsburgh, USA, pp. 233–236, 1994.

[226] D. L. Reilly, L. N. Cooper and C. Elbaum, "A Neural Model for Category Learning", Biological Cybernetics, Vol. 45, pp. 35–41, 1982.

[227] R. P. Lippmann, "An Introduction to Computing with Neural Nets", IEEE Acoustics, Speech and Signal Processing Magazine, Vol. 4, No. 2, pp. 4–22, 1987.

[228] S. Lee and R. M. Kil, "Multilayer Feedforward Potential Function Network", 2nd Int. Conf. on Neural Networks, Vol. 1, pp. 161–171, 1988.

[229] J. Moody and C. J. Darken, "Fast Learning in Networks of Locally-tuned Processing Units", Neural Computation, Vol. 1, pp. 281–294, 1989.

[230] R. P. Lippmann, "Pattern Classification Using Neural Networks", IEEE Communications Magazine, pp. 47–64, November 1989.

[231] B. Widrow and M. A. Lehr", 30 Years of Adaptive Neural Networks: Perceptron, Madaline and Backpropagation", Proc. of the IEEE, Vol. 78, pp. 1415–1442, 1990.

[232] T. Kohonen, "The Self-organizing Map", Proc. of the IEEE, Vol. 78, pp. 1464–1480, 1990.

[233] P. J. Werbos, "Backpropagation Through Time: What It Does and How to Do It", Proc. of the IEEE, Vol. 78, pp. 1550–1560, 1990.

[234] D. F. Specht, "Probabilistic Neural Networks", Neural Networks, Vol. 3, pp. 109–118, 1990.

[235] R. A. Jacobs, M. I. Jordan, S. J. Nowlan and G. E. Hinton, "Adaptive Mixtures of Local Experts", Neural Computation, Vol. 3, pp. 79–87, 1991.

[236] J. A. Leonard and M. A. Kramer, "Radial Basis Function Networks for Classifying Process Faults", IEEE Control Systems Magazine, Vol. 11, No. 3, pp. 31–38, 1991.

[237] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds and D. B. Rosen, "Fuzzy Artmap: A Neural Network Architecture for Incremental Supervised Learning of Analog Multidimensional Maps", IEEE Trans. on Neural Networks, Vol. 3, pp. 698–713, 1992.

[238] P. K. Simpson, "Fuzzy Min-max Neural Networks - Part 1: Classification", IEEE Trans. on Neural Networks, Vol. 3, pp. 776–786, 1992.

[239] G. D. Wilensky and N. Manukian, "The Projection Neural Network", International Joint Conference on Neural Networks, Baltimore, USA, Vol. 2, pp. 358–367, 1992.

[240] M. I. Jordan and R. A. Jacobs, "Hierarchies of Adaptive Experts", Advances in Neural Information Processing Systems, Vol. 4, pp. 985–992, 1992.

[241] C. M. Bishop, "Neural Networks and Their Applications", Review of Scientific Instruments, Vol. 65, pp. 1803–1832, 1994.

[242] S. Mitra and S. K. Pal, "Fuzzy Multi-layer Perceptron, Inferencing and Rule Generation", IEEE Trans. on Neural Networks, Vol. 6, pp. 51–63, 1995.

[243] M. A. Kramer and J. A. Leonard, "Diagnosis Using Back-propagation Neural Networks - Analysis and Criticism", Computers and Chemical Engineering, Vol. 14, pp. 1323- 1338, 1990.

[244] E. Parzen, "On Estimation of a Probability Density Function and Mode", Ann. Math. Statistics, Vol. 33, pp. 1065–1076, 1962.

[245] A. K. Jain and M. D. Ramaswami, "Classifier Design with Parzen Windows", in *Pattern Recognition and Artificial Intelligence*, eds. E. S. Gelsema, L. N. Kanal, Elsevier, pp. 211–228, 1988.

[246] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification", IEEE Trans. on Information Theory, pp. 21–27, 1967.

[247] W. S. Meisel, "Potential Functions in Mathematical Pattern Recognition", IEEE Trans. on Computers, Vol. 18, pp. 911–918, 1969.

[248] Y. Peng and J. A. Reggia, "A Probabilistic Causal Model for Diagnostic Problem Solving Part 2: Diagnostic Strategy", IEEE Trans. on Systems, Man and Cybernetics, Vol. 17, pp. 395–406, 1987.

[249] J. R. Josephson, B. Chandrasekaran, J. W. Smith and M. C. Tanner, "A Mechanism for Forming Composite Explanatory Hypothesis", IEEE Trans. on Systems, Man and Cybernetics, Vol. 17, pp. 445–454, 1987.

[250] T. Bylander, D. Allemang, M. C. Tanner and J. R. Josephson, "The Computational Complexity of Abduction", Artificial Intelligence, Vol. 49, pp. 25–60, 1991.

[251] J. R. Josephson, S. G. Josephson, *Abductive Inference*, Cambridge University Press, 1994.

[252] E. Binder, H. Egger, A. Hummer, M. Muhr and J. Schernthanner, "Predictive Maintenance of Generators", CIGRE, Paris, France, Paper 11–305, 1992.

[253] T. Kawamura, T. Suzuki, E. Zaima, I. Ohshima, T. Fujimoto, S. Kobayashi, T. Yamagiwa and K. Ibuki, "Development of GIS Diagnosis and On-site Testing", CIGRE, Paris, France, Paper 23/33–01, 1992.

[254] B. Florkowska and R. Wlodek, "Pulse Height Analysis of Partial Discharges in Air", IEEE Trans. on Electrical Insulation, Vol. 28, pp. 932–940, 1993.

[255] C. Hudon and R. H. Rehder, "Recognition of Phase Resolved Partial Discharge Patterns for Internal Discharges and External Corona Activity", 5th Int. Conf. on Conduction and Breakdown in Solid Dielectrics, Leicester, England, pp. 386–392, 1995.

[256] W. Y. Huang and R. P. Lippman, "Comparisions between Neural Net and Conventional Classifiers", 1st IEEE Int. Conf. on Neural Networks, San Diego, USA, pp. 485–493, 1987.

[257] W. Y. Huang and R. P. Lippman, "Neural Net and Traditional Classifiers", in *Neural Information Processing Systems*, ed. D. Z. Anderson, American Institute of Physics, pp. 387–396, 1988.

[258] G. T. Toussaint, "Bibliography on Estimation of Misclassification", IEEE Trans. on Information Theory, Vol. 20, pp. 472–479, 1974.

[259] P. A. Lachenbruch and M. R. Mickey, "Estimation of Error Rates in Discriminant Analysis", Technometrics, Vol. 10, pp. 1–11, 1968.

[260] M. Stone, "Cross-validatory Choice and Assessment of Statistical Predictions", Journal of Royal Statistical Society, Vol. B36, pp. 111–148, 1974.

[261] B. Efron, "Estimating the Error Rate of a Prediction Rule: Improvement on Cross-validation", Journal of the American Statistical Association., Vol. 78, pp. 316–331, 1983.

[262] A. K. Jain, R. C. Dubes and C.-C. Chen, "Bootstrap Techniques for Error Estimation", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 9, pp. 628–633, 1987.

[263] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley, 1992.

Manuscript was received on 11 April 1995, in revised form 24 August 1995.