

## Research Article

# Effectiveness of Partition and Graph Theoretic Clustering Algorithms for Multiple Source Partial Discharge Pattern Classification Using Probabilistic Neural Network and Its Adaptive Version: A Critique Based on Experimental Studies

S. Venkatesh,<sup>1</sup> S. Gopal,<sup>2</sup> and K. Kannan<sup>3</sup>

<sup>1</sup> Department of Electrical and Electronics Engineering, School of Electrical and Electronics Engineering, SASTRA University, Tirumalaisamudram, Tamil Nadu, Thanjavur 613 401, India

<sup>2</sup> W.S. Test Systems Limited, 27th km Bellary Road, Doddajalla Post, Karnataka, Bangalore 562 157, India

<sup>3</sup> School of Humanities and Sciences, SASTRA University, Tirumalaisamudram, Tamil Nadu, Thanjavur 613 401, India

Correspondence should be addressed to S. Venkatesh, venkatsri73in@gmail.com

Received 29 December 2011; Accepted 22 June 2012

Academic Editor: Raj Senani

Copyright © 2012 S. Venkatesh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Partial discharge (PD) is a major cause of failure of power apparatus and hence its measurement and analysis have emerged as a vital field in assessing the condition of the insulation system. Several efforts have been undertaken by researchers to classify PD pulses utilizing artificial intelligence techniques. Recently, the focus has shifted to the identification of multiple sources of PD since it is often encountered in real-time measurements. Studies have indicated that classification of multi-source PD becomes difficult with the degree of overlap and that several techniques such as mixed Weibull functions, neural networks, and wavelet transformation have been attempted with limited success. Since digital PD acquisition systems record data for a substantial period, the database becomes large, posing considerable difficulties during classification. This research work aims firstly at analyzing aspects concerning classification capability during the discrimination of multisource PD patterns. Secondly, it attempts at extending the previous work of the authors in utilizing the novel approach of probabilistic neural network versions for classifying moderate sets of PD sources to that of large sets. The third focus is on comparing the ability of partition-based algorithms, namely, the labelled (learning vector quantization) and unlabelled (*K*-means) versions, with that of a novel hypergraph-based clustering method in providing parsimonious sets of centers during classification.

## 1. Introduction

Among various techniques for insulation diagnosis, partial discharge (PD) measurement is considered a vital tool since it is inherently a nondestructive testing technique. PD is an electrical breakdown confined to a localized region of the insulating system of a power apparatus. PD, which may result in physical deterioration due to chemical degradation of the insulation system of a power apparatus, may occur as internal discharges in cavities, voids, blow-holes, gaps at the interfaces, and so forth or as external discharges on the surface imperfections, at sharp points and protrusions (corona discharges) etc. It is of major practical relevance for researchers and operators handling utilities to be able

to discriminate sources of PD, geometry, and location since such measurements are intimately related to the condition monitoring and diagnosis of the insulation system of such equipments. A few pertinent attributes [1] of PD pulses are their magnitude, rise time, recurrence rate, phase relationship of occurrence, time interval between successive pulses, discharge inception, and extinction voltage. Due to the advances in digital hardware systems, increase in the computational speed of processors and coprocessors and advancements in associated data acquisition systems there has been renewed focus among researchers in carrying out PD analysis [2]. More so, in recent years, the trend has shifted to recognition of patterns due to multiple sources of PD since these are often encountered during on-site, real time

measurements wherein distinguishing various sources of PD becomes increasingly challenging.

Diverse methodologies [3] have been adopted by several researchers to create a comprehensive and reliable system for discrimination and diagnosis of PD sources such as artificial neural network (ANN) [4–11], fuzzy logic controller (FLC) [12, 13], fractal features [14, 15], hidden Markov model [16–18], fast Fourier transform (FFT), and wavelet transform [19, 20] etc. Though attempts to classify single and partially overlapped sources of PD patterns have been successful to a fair degree [21], complexities in classifying fully overlapped patterns in practical insulation systems, complex non-Markovian characteristic of discharge patterns [22, 23], variation in the pulse patterns due to varying applied voltages in real time practical systems and so forth still continue to present substantial challenges [24].

Three major facets have been taken up for detailed study and analysis during the classification of multi-source PD patterns. The first aspect pertains to ascertaining the ability of the PNN versions without clustering algorithms in handling ill-conditioned and large training datasets, assessing the role of partition-based clustering algorithms (labelled: versions of LVQ algorithms and unlabelled: versions of  $K$ -Means algorithms) as compared to a novel graph theoretic based clustering techniques (hypergraph) in providing frugal sets of representative centers/during the training phase and analysis of the role played by the preprocessing/feature extraction techniques in addressing the curse of dimensionality and facilitating the classification task. In addition, a well-established estimation method that utilizes the inequality condition pertaining to various statistical measures of mean has been implemented as a part of the feature extraction technique to ascertain the capability of the proposed NNs in classifying the patterns. Further, exhaustive analysis is carried out to determine the role played by the free parameter (variance parameter) in distinguishing the classes, number of iterations, and its impact on computational cost during the training phase in NNs which utilize the clustering algorithms and the choice of the number of clusters/codebook vectors in classifying the patterns.

## 2. Preprocessing, Feature Extraction, and Neural Networks for Partial Discharge Pattern Classification: A Review

*2.1. Preprocessing and Feature Extraction.* A wide range of preprocessing and feature extraction approaches have been utilized by researchers worldwide for the task of PD pattern classification. Researchers involved in studies related to identification and discrimination of PD sources have usually resorted to the phase-resolved PD (PRPD) approach wherein methods based on statistical operators which include measures based on moments (skewness and kurtosis) [25–28], measures based on dispersion (range, standard deviation, variance, quartile deviation, etc.), central tendency (arithmetic mean, median, moving average etc.), cross-correlation, and discharge asymmetry and have been widely utilized. In studies related to time-resolved PD

analysis, pulse characteristic tools which include parameters such as pulse rise time, decay time, pulse width, repetition rate, quadratic rate, and peak discharge magnitude have also been attempted. Feature vectors consisting of average values of the spectral components in the frequency domain in analysis wherein signal-processing-related tools are utilized.

*2.2. Neural Networks for Pattern Recognition.* The prelude to PD pattern recognition studies can be traced to [29] wherein the multilayer perceptron-(MLP-) based feedforward neural network (FFNN) with back propagation algorithm (BPA) that has been attempted for training of the network was a remarkable success. Though the initial study was noteworthy and provided exciting avenues, further analysis pertaining to exhaustive data indicated that the basic version was computationally expensive due to long training epochs. Further studies with radial basis function (RBF) neural networks as reported in [30] showed improved performance and convergence during the supervised training phase with better discrimination of the decision surface of the feature vectors. However the tradeoff between unreasonably long training epochs and improved classification rate continued to present challenges to researchers.

Subsequently, unsupervised learning neural networks such as self-organized map (SOM), counter propagation NN (CPNN) [31], and adaptive resonance theory (ART) [32] have been utilized for classification of single-source PD signatures with a considerable level of satisfaction. However aspects such as complications related to the inherently non-Markovian nature of pulses further aggrandized by varying applied voltages during normal operation, apparently predictable incidence of ill-conditioned data obtained from modern digital PD measurement and acquisition systems which present considerable hurdles during large dataset training, and complexities in discriminating fully overlapped multisource PD signatures in practical insulation systems clearly substantiate on the need for a renewed focus on realizing a comprehensive yet simple NN scheme as a tool for the classification task.

Incidentally, the initial studies taken up earlier by the authors of this research in classifying small dataset PD patterns using PNN and its adaptive version [33, 34] clearly offer interesting solutions to difficulties related to large dataset training and classification in addition to providing a seemingly conceivable opportunity of utilizing a straightforward yet a reliable tool, since the PNN stems from a background based on sound theory related to statistics and probability. The standard version of the PNN (OPNN) and its adaptive version (APNN) are based on the strategy that combines utilizing a nonparametric density estimator (Parzen window) for obtaining the probability density estimates with that of a Bayesian classifier for decision making wherein the conditional density estimates are utilized for obtaining the class separability among the categories of the decision layer. It is pertinent to note that the only tunable part of the NN that requires to be tweaked for ensuring appropriate training is the variance (smoothing) parameter thus making the topology of the NN a plain yet a robust approach. It is evident, hence, that motivation

for this research is on ascertaining the capability of basic PNN versions (without and with clustering algorithms) in classifying multiple sources of PD at varying applied voltages. The effectiveness of these algorithms in to tackle large and ill-conditioned datasets acquired from the digital PD measurement and acquisition system which may lead to overfitting during the training phase is also studied.

### 3. Probabilistic Neural Network and Its Adaptive Version

PNN [35–37] is a classifier version based on “multivariate probability density estimation.” It is a model which utilizes the competitive learning strategy: a “winner-takes-all” attitude. The original (OPNN) and the adaptive versions of PNN (APNN) do not have feedback paths. PNN combines the Bayesian technique for decision-making with a nonparametric estimator (Parzen window) for obtaining the probability density function (PDF). The PNN network as described in Figure 1 consists of an input layer, two hidden layers (one each for exemplar and class layers), and an output layer.

Some of the merits of the PNN [38] include its ability in training with several orders of magnitude faster than the multilayer feedforward NN, capacity in providing mathematically credible confidence levels during decision making, inherent strength in handling the effects of outliers etc. One distinct disadvantage pertains to the need for large memory capability for fast classification. However, this aspect has been circumvented successfully in recent times since versions which have been implemented with appropriate modifications have been developed. Recently, the authors of this research have also successfully utilized a few variants of such modifications for multi-source PD pattern classification [39, 40].

Each exemplar node produces a dot product of the weight vector and the input sample, wherein the weights entering the node are from a particular sample. The product passes through a nonlinear activation function, that is,  $\exp[(x^T w_{ki} - 1)/\sigma^2]$ . The second hidden layer contains one summation unit for each class. Each summation (class) node receives the output from the pattern nodes associated with a given class given by  $\sum_{i=1}^{N_k} \exp[(x^T w_{ki} - 1)/\sigma^2]$ . The output layer has as many neurons as the number of categories (classes) considered during the study. The output nodes are binary neurons that produce the classification decision based on the condition  $\sum_{i=1}^{N_k} \exp[(x^T w_{ki} - 1)/\sigma^2] > \sum_{i=1}^{N_j} \exp[(x^T w_{ki} - 1)/\sigma^2]$ .

**3.1. Normalization Procedure in Modelling Pattern Unit.** The pattern unit in Figure 1 requires normalization of the input and exemplar vectors to unit length. A variety of normalization methods such as Euclidean, Minkowski (city block), and Mahalanobis may be utilized during the NN implementation, though the most popular being the Euclidean and the city block norms. Figure 2 can be made independent of the requirement of unit normalization by adding the length of both vectors as inputs to the pattern unit.

A basic variant of the PNN called the adaptive PNN (APNN) [41, 42] offers a viable mechanism to vary the free parameter “ $\sigma$ ” (variance parameter) or the smoothing parameter within a particular category (class node). While the OPNN utilizes a common value for all of the classes, the APNN employs different values of  $\sigma$  for each class based on computing the average distance  $\sigma = g \cdot d_{ave}$  from Euclidean distances among various feature vectors while “ $g$ ” is a constant which necessitates adjustment. An additional aspect of this approach is that a simplified formula of probability density function (PDF) is used which obviates the necessity for normalization and hence a considerable amount of computation is reduced.

### 4. Partitioning and Graph Theoretic Clustering Algorithms: An Overview

Clustering deals with segregating a set of data points into nonoverlapping groups or cluster points wherein the points in the group are “more similar” to one another than to points in other groups [43]. The term “more similar” when used to clustered points, usually refers to closeness by a credible quantification of proximity. When a dataset is clustered, each point is allocated to a particular cluster and every cluster can be characterized by a single reference point usually an average of the points in the cluster. A wide range of clustering algorithms has been utilized by researchers in diverse engineering applications which fall under eight major categories [44]. These are based on similarity and sequence similarity measures, hierarchy, square error measures, mixture density estimation, combinatorial search, kernel, and graph theory. While the hierarchical clustering groups data with sequence of partitions from solitary cluster to a cluster including all clusters, partition clustering on the other hand divides data objects into prefixed clusters without the hierarchical composition. Partition-based clustering methods include square error; density estimate includes vector quantization,  $K$ -Means, and expectation maximization (EM) with maximum likelihood (ML).

Any specific segregation of all points in a dataset cluster is called “partitioning”. Data reduction is accomplished by replacing the coordinates of each point in a cluster with the coordinates of the appropriate reference point. The effectiveness of a particular clustering method depends on how closely the reference points represent the data as well as how fast the algorithm proceeds and gets processed. If the data points are tightly clustered around the centroid, the centroid will be representative of all the points in that cluster. The standard measure of the spread of a group of points about its mean is the variance or the sum of the square of the distance between each point and the mean. If the data points are close to the mean, the variance will be small. The level of error “ $E$ ” as a measure indicates the overall spread of data points about their reference points. To achieve a representative clustering,  $E$  should be as small as possible. When clustering is done for the purpose of data reduction, the goal is not in finding the best partitioning but rather a reasonable consolidation of “ $N$ ” data points into “ $k$ ”

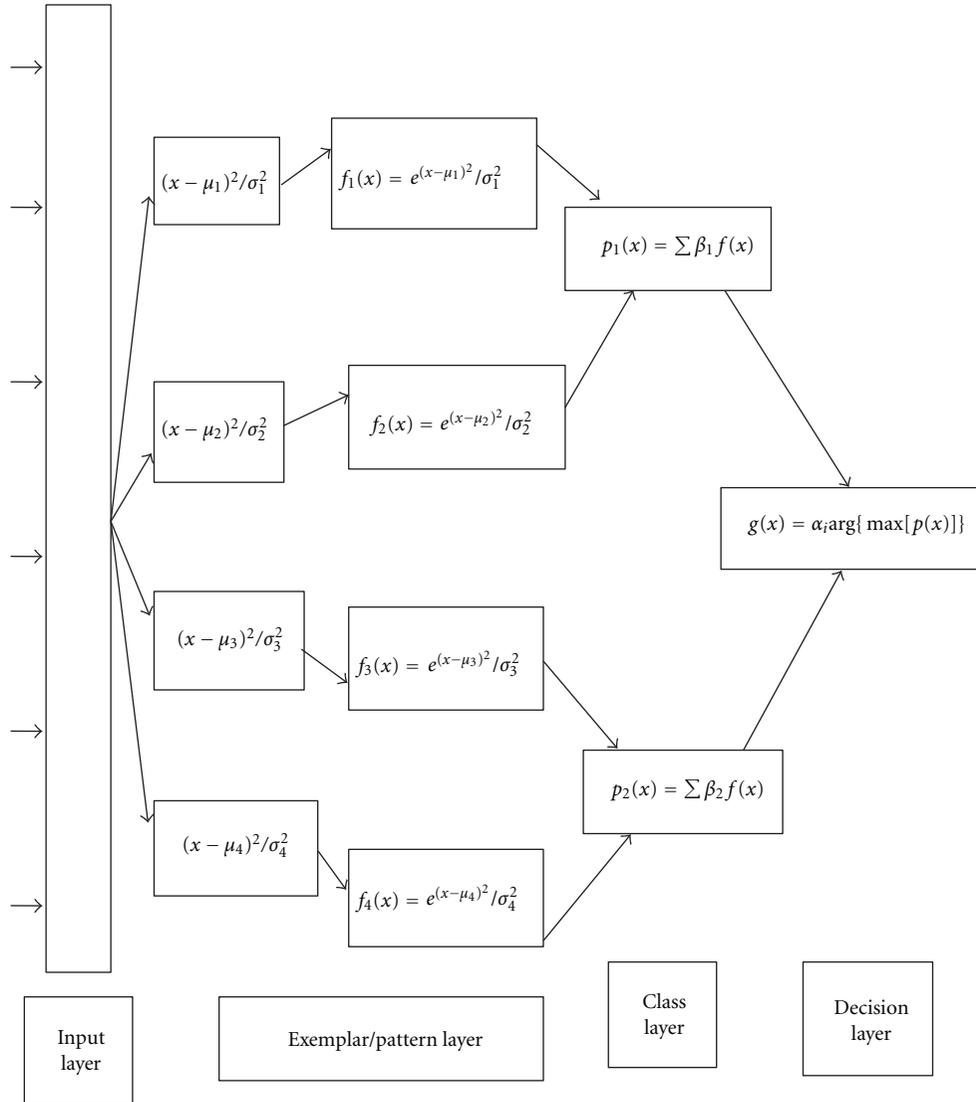


FIGURE 1: Architecture of probabilistic neural network.

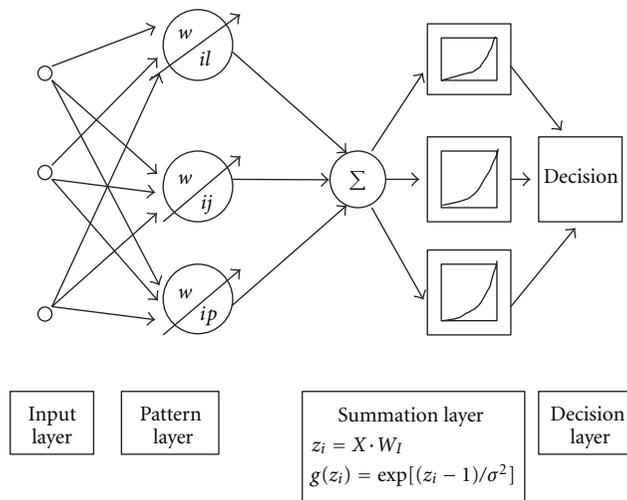


FIGURE 2: Normalization in a pattern unit: original PNN.

clusters and if possible some efficient means to improve the quality of the initial partitioning. In this aspect a family of iterative-partitioning algorithms either of labelled or unlabelled versions has been developed by researchers. Over the years several clustering algorithms have been proposed by researchers which include the hierarchical clustering (agglomerative, stepwise optimal), online clustering (leader-follower clustering), and graph theoretic clustering.

Though the graph theoretic representation of data may also provide avenues for clustering, its limitation from the viewpoint of complex applications stems from the fact that it utilizes binary relations which may not comprehensively represent structural properties of temporal data, the nature of association being binary neighbourhood. In this context it is worth noting that only recently, hypergraph (HG) theory and its relevant properties have been exploited by researchers for designing computationally compact algorithms for pre-processing data in various engineering applications such as image processing and bioinformatics etc [45] due to its inherent strength in representing data based on both topological and geometrical aspects while most other algorithms are topology based only. Hypergraph (HG) deals with finite combinatorial sets and has the ability to capture both topology and geometrical relationships among data.

Hence, it is apparent from this discussion that the choice of the appropriate type of clustering technique would play a vital role in handling the classification of large dataset PD.

**4.1. Labelled Partition-Based Clustering Learning Vector Quantization Versions.** Kohonen's [46] learning vector quantization (LVQ) is basically a pattern-classification-supervised learning version wherein each output neuron represents a particular class/category. The weight vector for an output neuron is usually called as a reference (codebook) vector of the class that the unit signifies. During training, the output units are placed by adjusting the weight vector to approximate the decision hypersurface of the Bayesian classifier. During testing of the PNN and its adaptive version using LVQ clustering technique [47], the LVQ classifies an input vector by assigning it to the same class as the output unit which has its weight vector the closest.

**4.1.1. LVQ1.** This simple algorithm proposes updating the weight towards the new input vector ( $x_i$ ) if the input and the weight vector belong to the same class or updating the weight away from the input if the input and the weight vector belong to different classes (determined by finding the output pertaining to minimum distance, i.e.,  $\|x_i - w_j\|$ ).

**4.1.2. LVQ2.** The modification in this version relates to updating the weights for the runner up distance based on the constraint that the ratios of runner up ( $d_r$ ) and closest distance ( $d_c$ ); that is,  $d_r/d_c > (1 - \epsilon)$  and  $d_c/d_r < (1 + \epsilon)$  ( $\epsilon$  is the window describing the error in the variance) in addition to the restriction that the distance between  $x_i$  and codebook belongs to two different classes for closest and runner up distance and that  $x_i$  belongs to codebook whose target is runner up. When both the closest and next closest distance are not the target output, updating of  $d_r$  and  $d_c$  is

swapped. When the target is the nearest codebook, then the updating of weight for that particular exemplar is not carried out.

**4.1.3. LVQ3.** Additional enhancements on the previous versions enable the learning of two closest vectors which satisfy the window condition  $\min(d_{c1}/d_{r2}, d_{c2}/d_{r1}) > (1 - \epsilon)(1 + \epsilon)$ . In such a case the weights are updated as  $y_c(t + 1) = y_c(t) + \beta(t)[x(t) - y(t)]$  for both  $y_{c1}$  and  $y_{c2}$ . The learning rate  $\beta(t)$  is a multiple of learning rate  $\alpha(t)$ , and its typical value ranges between 0.1 and 0.5 with smaller values corresponding to a narrower window.

**4.2. Unlabelled Partition Based Clustering: K-Means Algorithm Versions.** K-means algorithm [48] locates and obtains the "c" mean (cluster center) vectors ( $\mu_1, \mu_2, \mu_3, \dots, \mu_c$ ). This rudimentary unlabelled clustering algorithm is called K-means algorithm commonly referred to as Lloyd (or) Forgy's K-Means. In order to facilitate in having better sets of cluster representatives and for ensuring a reasonable choice of the initial seed vector, various variants have been developed which include McQueen K-means, standard K-means, continuous K-means, and fuzzy K-means etc. to provide a better choice on the initial seed and consequently better sets of cluster representatives.

**4.2.1. Forgy's K-Means.** The algorithm describing this method is illustrated in Figure 3.

**4.2.2. Standard K-Means.** The distinct distinction from the Forgy K-means is in its more appropriate utility of the data at each step. Though the basic process for both algorithms is similar in the context of choice of the reference points and in the allocation of clusters to all data points, then using the cluster centroids as reference points in subsequent partitioning the distinctness is in nature of adjusting the centroids both during and after each partitioning. For a data "x" in cluster "i" if the centroid  $z_i$  is the nearest reference point, then adjustments are not carried out and the algorithm proceeds to the next sample data. On the other hand, if the centroid  $z_j$  of the cluster "j" is the reference point closest to the data "x," then x is reassigned to cluster j in addition to recomputing the centroids of the "losing" cluster "i" (minus point x) and the "gaining" cluster (plus point x) and moving the reference points  $z_i$  and  $z_j$  to the fresh centroids.

**4.3. Graph Theoretic Clustering Algorithm: Hypergraph.** A HG [49] "H" is a pair  $(X, \xi)$  consisting of a nonempty set X together with a family  $\bigcup_{i \in I} E_i = X, I = \{1, 2, \dots, n\}, n \in N$ . Figure 4 shows a generic HG representation.

An important structure that can be studied in a HG is the notion of an intersecting family. An intersecting family of hyperedges of a HG "H" is a family of edges of H which have pairwise nonempty intersections. There are two types of interesting families: (1) intersecting families with an empty intersection and (2) intersecting families with a nonempty intersection. A HG has the Helly property if each family of pair wise intersecting hyperedges has a nonempty

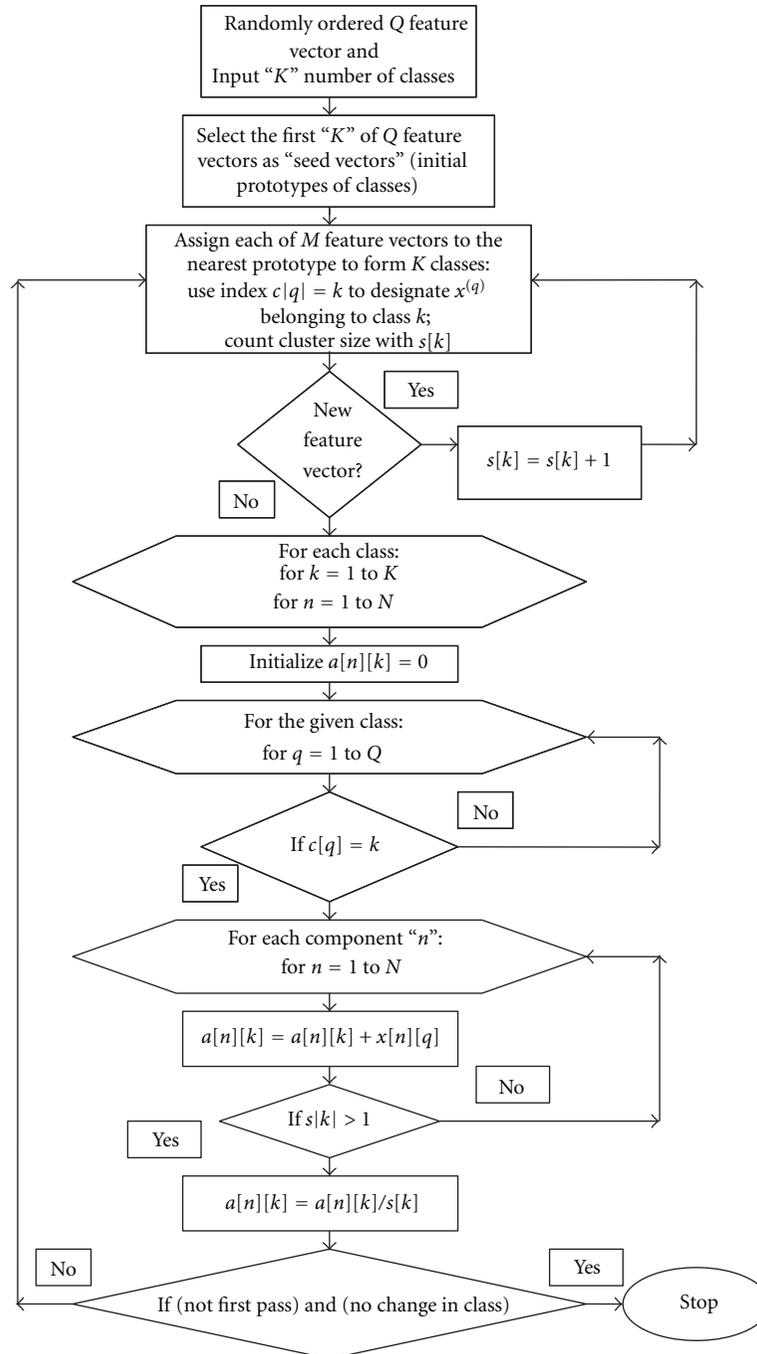


FIGURE 3: Flow chart of K-means clustering algorithm.

intersection (i.e., they belong to a star). Figure 5, represents two types of intersecting hyper-edges.

Several researchers in allied fields [50, 51] of engineering have utilized a variety of properties of HG such as the Helly, transversal, mosaic, and conformal for obtaining clustering algorithms pertaining to a diverse set of applications. The neighbourhood HG representation utilizes the Helly which plays a vital role in identifying homogeneous regions in the data augurs as well as serves as the main aspect for developing segmentation and clustering algorithms.

In the case of studies based on HG-based clustering and classification, the preprocessed data obtained as discussed in Section 6 is represented as  $V_i = (\varphi_i, q_i, n_i)$ ,  $i = 1, 2, 3, 4, \dots, m$  where "m" is the number of vertices of the data per cycle. The data is grouped in terms of feature vectors which act as the best representatives of entire database. Hence if pair-wise intersecting edges are created from the entire data base, the Helly property of HG can be invoked to find the common intersection which in turn provides the feature vectors that represent the centers of a particular set

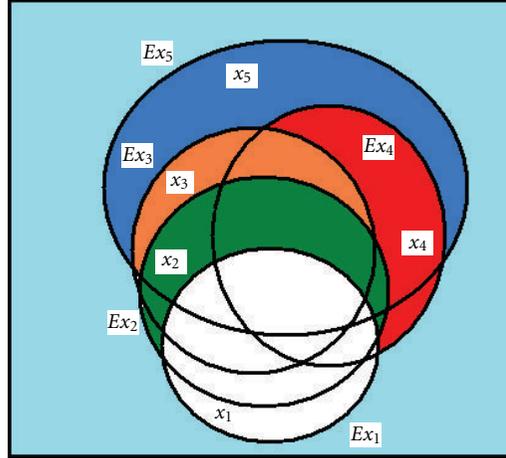
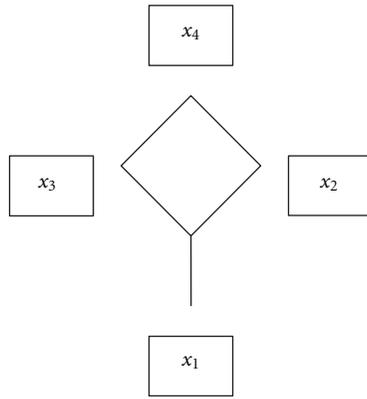


FIGURE 4: Generic representation of graph and HG.

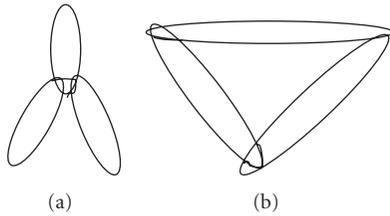


FIGURE 5: Representation of pairwise intersecting hyperedges with (a) nonempty and (b) empty intersection.

of data pertaining to the source of PD. Hence, a minimum distance metric scheme (Euclidean) is developed to obtain the nearest among various intersections of the intracluster and intercluster dataset so as to obtain the optimal set of common intersection vectors that serve as the centers representing the dataset. These feature vectors are taken as training vectors of the PNN.

## 5. Partial Discharge: Laboratory Setup, Artificially Simulated Benchmark Models, and Data Acquisition

**5.1. PD Laboratory Test Setup.** Comprehensive studies pertaining to single- and multi-source PD pattern recognition have been carried out using a W.S. Test Systems Make (Model no.: DTM-D) digital PD measurement system suitable for measuring PD in the range 2–5000 pC with Tektronix built-in oscilloscope (TDS 2002B) provided with a tunable filter-insert (Model: DFT-1) with a selectable center frequency in the range of 600 kHz–2400 kHz at a bandwidth of 9 kHz. PD pulses acquired from the analogue output terminal are exhibited on the built-in oscilloscope. The measured partial discharge intensity is displayed in picocoulomb (pC).

PDGold software developed by HV Solution UK is interfaced with the PD measurement system to acquire the PD patterns. Window gating facility is provided by the PD acquisition system to suppress background noise.

The test setup and various stipulations of the test procedure comply with IEC 60 270 [52]. Further, in order to improve the transfer characteristics of the test system, a 1 nF coupling capacitor is integrated to the test setup. An electronic reference calibrator (Model: PDG) ensures appropriate resolution of pulses during measurement and data acquisition. The straight detection and measurement test setup as recommended in IEC is utilized in carrying out the test. Figures 6, 7, and 8 show the test arrangement for PD measurement and acquisition system.

**5.2. Artificially Simulated Laboratory Benchmark Models for PD Pattern Classification.** Five categories of laboratory benchmark models have been fabricated to simulate distinct classes of single and multiple PD sources, namely, electrode bounded cavity, air corona, oil corona and electrode bounded cavity with air corona which would in turn serve as a validation technique to replicate the reference patterns as recommended in [53]. Internal discharges are simulated by an electrode bounded cavity of dimension 1 mm diameter and 1.5 mm depth on 12 mm thick polymethyl methacrylate (PPMA) of diameter 80 mm as shown in Figure 9. One category of external discharge (surface discharge) is simulated with 12 mm thick Perspex of 80 mm diameter as indicated in Figure 10. A second category of external discharge called the air corona discharge is replicated by an electrode of apex angle 85° attached to the high voltage terminal as shown in Figure 11. Corona discharge in oil is produced with a similar arrangement immersed in transformer oil as shown in Figure 12. Electrode bounded cavity with air corona is produced by inserting a needle configuration (2 mm) from the HV terminal in addition to a 2 mm bounded cavity in Perspex at the high voltage electrode as replicated in Figure 13.

**5.3. PD Signature and Pattern Acquisition System.** PD Gold is a data acquisition software which provides a system to acquire high-resolution PD signals at a high-sampling rate (1 sample per 2.5 nanoseconds). The system detects PD on

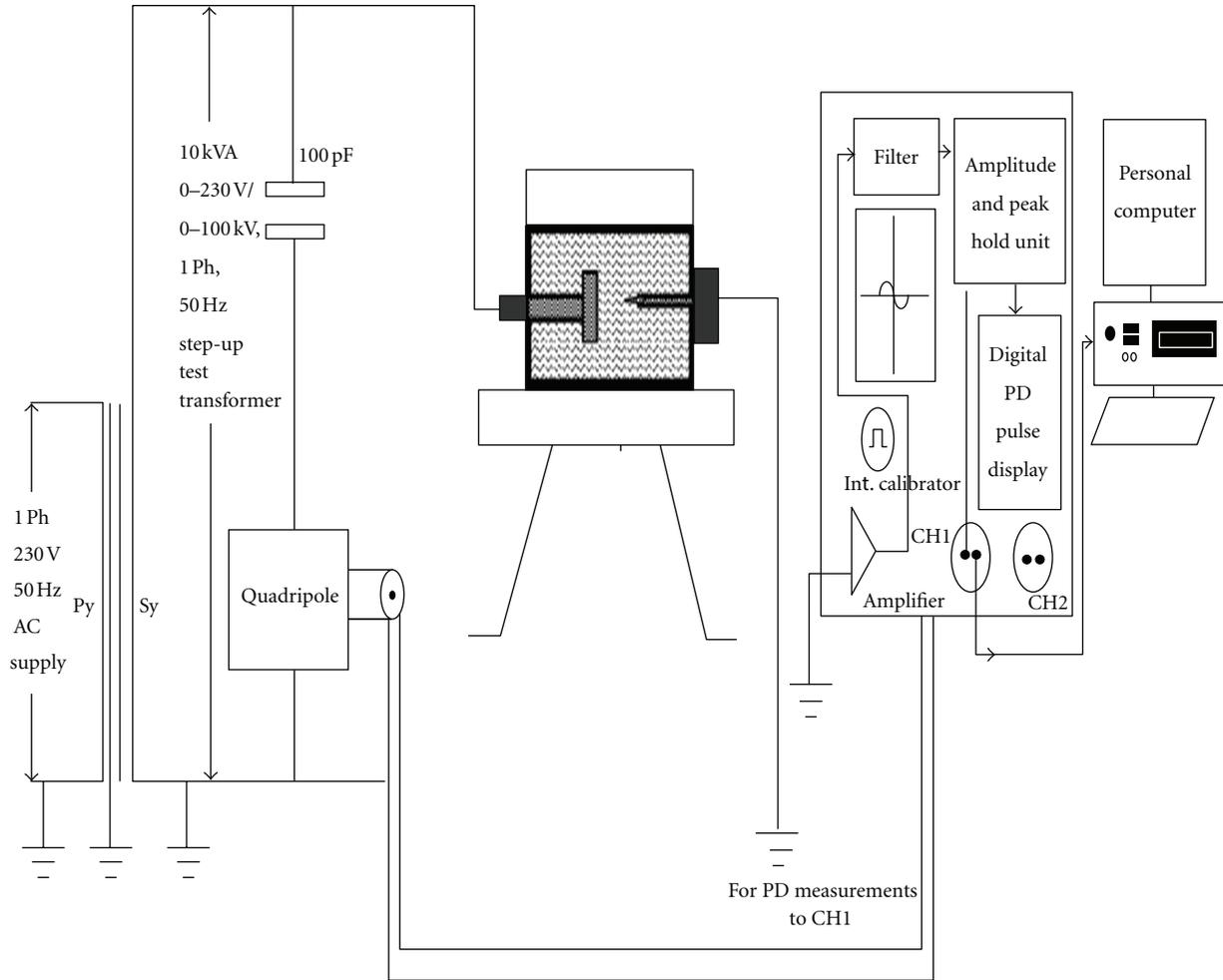


FIGURE 6: Typical laboratory test setup for PD pattern recognition studies.

a 50 Hz power cycle base thus enabling display of PD pulses in sinusoidal or elliptical forms usable in either auto or manual mode which in turn enables the user to observe the shape of the PD pulses detected and representing the PRPD patterns in real time. In the manual approach, the user has the facility to record the data for a considerable duration (in this study 5–15 minutes) which is acquired from a minimum of 240 to a maximum of 750 waveforms per channel.

Incidentally, for carrying out PD testing which would ensure credible acquisition of data it is essential to acquire fingerprints of PD signals under well-defined conditions. Hence, before testing, the test specimen is preconditioned in line with the requirements of the relevant technical committee. Since methods of cleaning and conditioning test specimens play a vital role during acquisition of the test data, preconditioning procedures indicated in [54] are adopted.

It is observed during exhaustive studies that for discharge sources listed in Tables 1 and 2, a time period of 5 minutes is usually sufficient to capture the inherent characteristics of PD. Figures 14 and 15 show typical PD pulses acquired during the testing, measurement, and acquisition process.

TABLE 1: Moderate dataset of PD laboratory models.

PD category	Type of PD	Label for classification	No. of PD patterns
(1)	Electrode bounded cavity	EC	10
(2)	Surface discharge	SD	10
(3)	Oil corona	OC	10
(4)	Air corona	AC	6
(5)	Electrode bounded cavity with air-Corona	ECAC	10
(6)	Electrode bounded cavity with surface discharge	ECSD	10

## 6. Preprocessing and Feature Extraction

For carrying out extensive training and testing of the PNN versions, the raw data is preprocessed in order to ensure compactness without compromising on unique details of the characteristic input feature vector. The significance of utilizing a wide variety of preprocessing methods is to enable in ascertaining the performance of the proposed NNs so

TABLE 2: Large dataset PD database of laboratory models with varying applied voltages.

Category of PD	Label for identification/classification	Type of PD	Applied voltage (kV)	Total no. of training patterns	Total no. of testing patterns
(1)	EC	Electrode bounded cavity	7.3 9.1 9.6	90	120
(2)	AC	Air corona	14 21 23	90	120
(3)	OC	Oil corona	21 29.1 32	90	120
(4)	ECAC	Electrode bounded cavity with air corona	7.3 9.1 10	90	120



FIGURE 7: Laboratory experimental test setup indicating the direct detection PD measurement methodology.

that tangible decisions may be taken on the role played by the various key parameters of the neural networks such as smoothing parameter, effect of outliers, and curse of dimensionality. The input data presented to the PNN is based on phase window technique wherein only simple statistical operators, namely, (1) measures based on maximum values of  $q$  ( $10^\circ$  and  $30^\circ$ ); (2) measures based on minimum values of  $q$  ( $10^\circ$  and  $30^\circ$ ); and (3) measures based on central tendency ( $10^\circ$  and  $30^\circ$ ), are utilized to ascertain the capability of the proposed PNN versions in classifying patterns as a preliminary case study. The authors of this research work have carried out earlier exhaustive studies based on the traditional statistical operators albeit with other NNs. Thus the major focus of this research is on assessing the capability of PNN algorithms in classifying multiple sources of PD utilizing both clustering algorithms and algorithms without the influence of clustering and its role in distinguishing the classes appropriately with parsimonious sets of centers.

Further, a new method of utilizing the inequality [55] given by harmonic mean (HM)  $\leq$  geometric mean (GM)  $\leq$  arithmetic mean (AM)  $\leq$  root mean square-based on

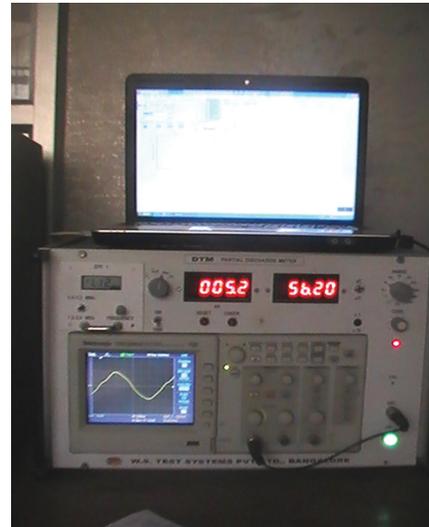


FIGURE 8: Digital PD measurement and acquisition system (DTM-D model).

measures of various types of mean utilized successfully by a few researchers in the field of target recognition serves as an effective yet simple technique in reducing the dimensionality of the input feature vector space. Hence, it has also been adapted in this research work to ascertain its effectiveness in providing a compact set of extracted features.

The acquisition of raw PD dataset was carried out as deliberated in Section 5, preliminarily for moderate set of multiple source PD patterns and subsequently for large datasets for single and multiple PD sources. The first studies are conducted for dataset consisting of a total of two sets of training database, that is, 20 and 25 sets. A total of 56 PD fingerprints samples were collected from 6 samples of benchmark models described in Section 5 of which 10 patterns are due to internal discharge (electrode bounded cavity), 10 pertain to oil corona, 10 patterns correspond to surface discharge, 6 fingerprints belong to air corona patterns and 10 patterns belong to electrode bounded cavity with air

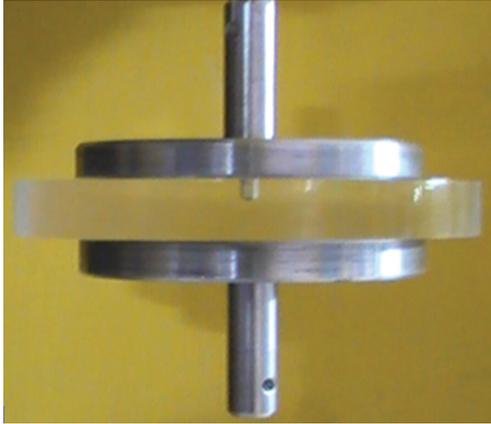


FIGURE 9: Laboratory model replicating electrode bounded cavity discharge.

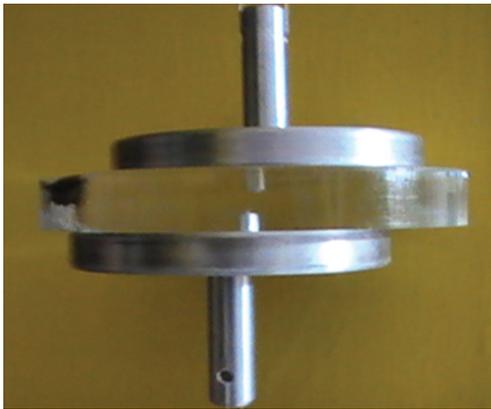


FIGURE 10: Model simulating surface discharge with electrode bounded cavity.

corona (multisource PD). The database obtained is indicated in Table 1.

The second analysis pertains to PD signatures for large dataset patterns acquired from the laboratory testing of 4 models simulating sources of PD. The total number of fingerprints in the database comprises ninety patterns of each type of the defect with thirty samples pertaining to each of the various applied voltages. It is to be noted that these patterns have been acquired online wherein the statistical variations in the pulse patterns for each cycle of the sinusoidal voltage exhibits the inherent non-Markovian nature, thus making the task during classification more difficult. The task becomes even more demanding due to different applied voltages which make the process of classification of pulse patterns complex. Rigorous study and analysis on the classification capability of the proposed NN is carried out for only one applied voltage for each category of PD. However, the limitations and aspects related to complexities in classifying large dataset due to varying applied voltages are also summarized. Table 2 shows the patterns acquired for large dataset from various sources of PD.



FIGURE 11: Laboratory model simulating air corona discharge with point configuration as high voltage electrode at an  $85^\circ$  apex angle.



FIGURE 12: Laboratory model simulating oil corona discharge with point configuration as high voltage electrode at an  $85^\circ$  apex angle.

It is pertinent to note from Table 2 that only eighteen sets (20% of the training dataset) pertaining to each source of PD (referred to as prototype/codebook vectors in the case of labelled clustering or random cluster centers in the case of unlabelled clustering) were taken up for finding the centers since it has been observed from our study that these representative fingerprints were sufficient for obtaining considerable number of centers which led to reasonable classification capability of PNN versions. This is notwithstanding the fact that NN literature studies have indicated the usual practice of using at least 50% as representative samples for the training phase though it would ideally be suitable to have two-thirds as the basis for training the NNs. However, further studies were taken up by the authors with 40% of codebook vectors for obtaining centers. It was made evident that enhanced classification capability was evinced by the NNs.

## 7. Neural Network Verification

The most prevalent verification methods, namely, the alphabet character recognition and the Fischer's Iris Plant database [56], are used for training and testing of the PNN versions for ascertaining the performance of the proposed PNN versions. Coding for versions of PNNs is developed using MATLAB 6.1, Release 12. The ability of

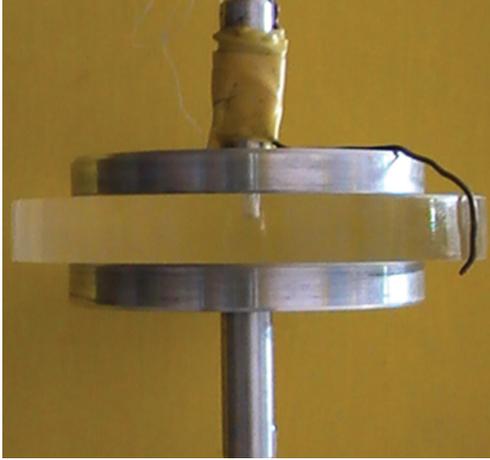


FIGURE 13: Laboratory model replicating electrode bounded cavity discharges overlapped on air corona discharges (multiple source discharges).

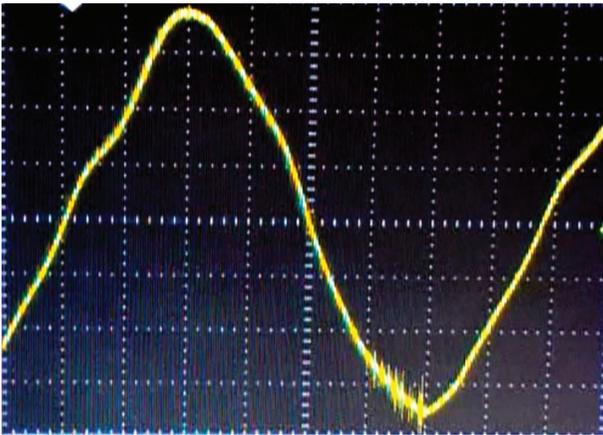


FIGURE 14: Typical waveform representation on the oscilloscope depicting air corona discharges on sinusoidal base.

the clustering algorithms and hence the number of codebook reference vectors/centroids as appropriate to the type of clustering formed have also been studied and found to be reasonably precise in classifying the divergent input vectors.

## 8. Analysis and Inferences

**8.1. Case Study 1: Discrimination Capability of OPNN and APNN without Clustering Algorithm for Moderate PD Datasets.** Based on the training and testing of PNN and its adaptive version with two sets of training data which include overlapped and single PD source patterns comprising 4 sets (3 nos. of single PD source and 1 number of void corona overlapped) and 5 sets (3 single PD sources and 2 numbers void corona and void-surface discharge overlapped), extensive observations and analysis are summarized.

TABLE 3: Classification capability of PNN and APNN for moderate database—without clustering algorithm.

Pre-processing Technique	Misclassifications in OPNN	Misclassifications in APNN
$\Phi-q_{\max}-n$ ( $30^\circ$ )	4 types—7 numbers (AC2, AC7, EC1AC2, EC2AC7, EC5AC2, EC6AC7)	4 types—5 numbers (EC1, EC5, SD6, SD9)
	5 types—15 numbers (AC2, AC7, EC1AC2, EC2AC7, EC3AC8, EC4AC9, EC2SD2, EC3SD3, EC4SD4, EC1, EC6AC7, EC6SD6, EC7SD7, EC8SD8)	5 types—6 numbers (EC1, EC6AC7, EC6SD6, EC7AC8, EC7SD7, EC8SD8)
$\Phi-q_{\min}-n$ ( $30^\circ$ )	4 types—8 numbers (EC5AC2, EC3AC8, EC2AC7, EC1AC2, EC6AC7, EC7AC8, EC8AC9)	4 types—5 numbers (EC1, EC6AC7, EC7AC8, EC8AC9)
	5 types—15 numbers (EC1AC2, EC2AC7, EC3AC8, EC5AC2, EC1SD1, EC2SD2, EC3SD3, EC4SD4, EC5SD5, EC1, EC6AC7, EC6SD6, EC7SD7, EC8SD8)	5 types—7 numbers (EC1, EC6AC7, EC6SD6, EC7AC8, EC7SD7, EC8SD8)
$\Phi-q_{\max}-n$ ( $10^\circ$ )	4 types—7 numbers (AC2, AC7, EC1AC2, EC5AC2, EC6AC7, EC8AC9)	4 types—4 numbers (EC2, EC3, EC5, SD6)
	5 types—8 numbers (EC5SD5, EC4SD4, EC1, EC6AC7, EC7AC8, EC8AC9, EC7SD7)	5 types—4 numbers (EC1, EC5, SD6, SD9)
$\Phi-q_{\min}-n$ ( $10^\circ$ )	4 types—8 numbers (AC2, AC7, EC3AC7, EC6AC2, EC8AC9, EC3AC8, EC2, EC4, EC6)	4 types—6 numbers (S6, S7, S10, V6C2, V8C9)
	5 types—12 numbers (AC2, AC7, EC3AC7, EC1AC2, EC6SD6, EC7SD7, EC8SD8, EC6AC2, EC2, EC4)	5 types—5 numbers (EC7SD7, EC6AC2, SD10, SD7, SD6)

### 8.1.1. Analysis of the Performance of OPNN

- (1) Since the basic version of PNN is an unsupervised learning scheme (without feedback for learning), the exemplar nodes are themselves the weight vector and hence these are not updated during the training phase (training phase is not a part of the rudimentary scheme). Hence, it is obvious that for effective learning higher number of exemplar nodes which are representative of the category of PD source during training would ensure enhancing

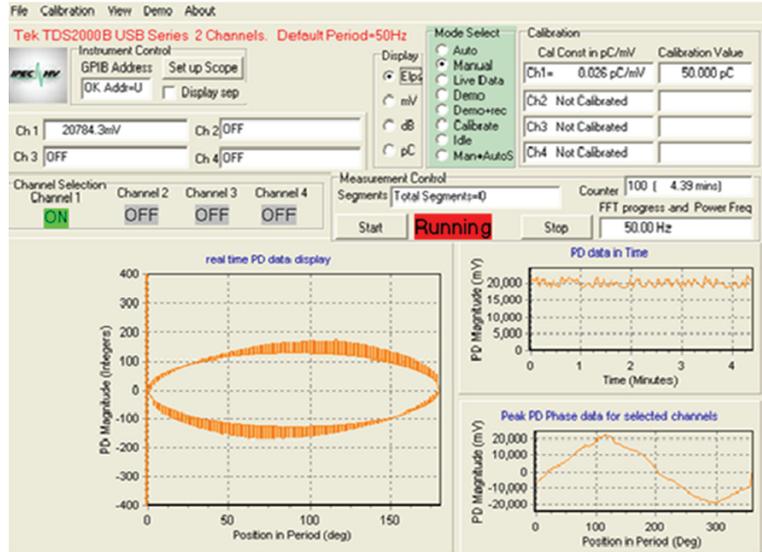


FIGURE 15: Typical sample of laboratory model testing of electrode bounded cavity with air corona PD acquired from the PD measurement and acquisition system.

the classification capability of both versions of PNN. Though a minor variation in classification capability of the PNN version may be obtained by tweaking the variance parameter since the focus of the research is on comparing the characteristics of clustering algorithms, a fixed value of smoothing parameter is taken for the purpose of analysis during classification. The classification capability is summarized in Table 3.

- (2) Since it is also made evident during detailed study that issues related to overfitting would be an important aspect while training large non-Markovian PD datasets, this algorithm suffers from the drawback of requirement of large memory during the training phase.

8.1.2. Analysis on the Performance of APNN

- (1) It is also evinced from detailed study that since the adaptive version provides a mechanism for having independent variance parameter for unique class labels, this version in almost all cases learnt well during the training phase (though this network structure also does not include supervised learning). This feature is evident from the modifications made in the structure of the APNN (due to the separate values of the variance parameter pertaining to each class decision boundaries). Table 3 and Figure 16 substantiate this aspect.
- (2) Nevertheless, since the basic variant of PNN also does not involve training and supervision during learning, considerable numbers of misclassifications are noticed, more so, pertaining to fully overlapped multi-source (electrode bounded cavity with surface discharge) PD signatures. The difficulties during classification of such overlapped signatures are evident from the nature of also from the nature of

Classification capability of OPNN- and APNN-multiple PD sources

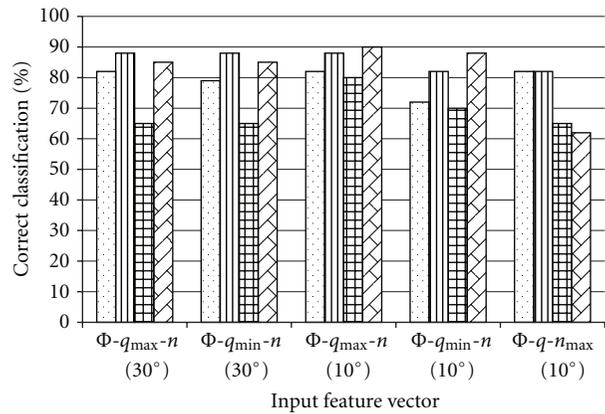


FIGURE 16: Classification capability of OPNN and APNN with five types of feature inputs with 4- and 5-types of overlapped patterns—without clustering algorithm (in the histogram dotted chequered blocks refer to 4 and 5 type inputs to PNN; striped and brick blocks refer to 4- and 5-type inputs to APNN).

hyperboundary separation, wherein values of the smoothing parameter are indicated in Table 5.

8.2. Case Study 2: Performance of OPNN and APNN with Labelled (LVQ Versions) Algorithms for Moderate PD Datasets

- (1) Fewer misclassifications are noticed during training of multiple source PD patterns in most of the LVQ variants considered for study in this research. The only exception noticed is with the measures based on minimum and maximum values wherein considerable number of misclassification are

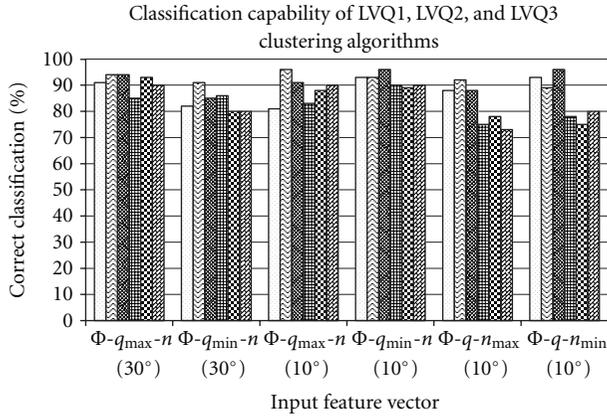


FIGURE 17: Classification capability of OPNN and APNN with six types of feature inputs with 4- and 5-types of overlapped patterns—with LVQ clustering algorithms (in the histogram dotted, chequered blocks refer to 4- and 5-type inputs to PNN; striped and brick blocks refer to 4- and 5-type inputs to APNN).

observed for fully overlapped PD source considered in this study. Results of the comprehensive set of studies are shown in Table 4 and Figure 17.

- (2) It is also of considerable importance to note from Table 5 that the decision hyperboundaries that separate the various categories of PD sources are found to be very sharp (small values of the variance parameter). This clearly indicates that the complexities pertaining to classification of multi-source PD signatures in addition plausible inconsistencies during data acquisition for subsequent training and testing by PNN variants.
- (3) Another prominent feature made evident from Table 5 is the similarity in the range of values of variance parameter for various categories of PD sources. Incidentally the values of variance parameter in the case of APNN are found to be almost similar thus signifying the similar nature of both Bayesian-based strategies in creating hypersurface boundaries. The performance of PNN versions which utilize the variants of LVQ algorithms is summarized in Figure 17.

### 8.3. Case Study 3: Role of the Trainable Part in Unsupervised and Supervised PNN Versions

- (1) It is pertinent to note from Table 5 that in the case of all the versions of LVQ clustering-(LVQ1, 2, and 3) based PNNs, the range of the variance parameter is between 0.01 and 0.05 which describe the feature for void defect. Similarly the value of  $\sigma_4$ , that is, void-corona overlapped pattern, is also reasonably similar but for one specific case with LVQ3 only. This establishes the fact, already stated by researchers in identifying and classifying the overlapped void-corona patterns. In addition, from the viewpoint of decision of the boundary hyperplane, considerable clarity in separation of class boundaries is noticed.

- (2) However, in the case of void-surface overlapped patterns the value of variance is considerably divergent in various versions of LVQ. This is vividly observed in the case of input feature vector using measures based on minimum and maximum values of number of pulses.
- (3) Since it is noticed that the value of variance parameter is narrow (peaked), it is evident that such a technique may not be appropriate for further fine tuning of the trained vectors. This technique might augur well only for large training datasets wherein wider class identification is expected thus possibly suggesting the need for more training for obtaining enough number of representative codebook vectors pertaining to a class for better class discrimination.

### 8.4. Case Study 4: Performance of OPNN and APNN for Large Dataset with Traditional Statistical Operators and Inequality Measures of Mean with Labelled (LVQ Versions) Algorithms

- (1) It is worth noting that the LVQ versions of algorithms are able to create a reasonably good parsimonious set of centers relevant to the four classes even with about 20% (6 codebook vectors for every 30 training datasets of each applied voltage) of prototype vectors. In this context, it is to be emphasised that these codebook vectors become the weight (centers/centroids) vectors which are now the representatives of the samples. Table 6 summarizes the classification capability of the LVQ-PNN variants.
- (2) It is also evident from Table 6 the superiority of the LVQ 2 version as a clustering algorithm for large dataset training as compared to the other types. This characteristic noticed in the course of this study by the authors of the research work has also been concurred by a few researchers in other allied areas of engineering [57].
- (3) When the study was extended to that of doubling of the number of reference vectors during training, the improved classification rate is noticed (about 90–95%) for almost all categories and types of preprocessing schemes of varying levels of compactness.
- (4) A perceptible difference in the classification capability of patterns pertaining to the feature extraction scheme that utilizes the inequality relation based on the measures related to the types of mean values (with both  $30^\circ$  and  $10^\circ$  phase window input features) has been observed.

### 8.5. Case Study 5: Performance of OPNN and APNN for Large Dataset with Traditional Statistical Operators and Inequality Measures of Mean with Unlabelled (K-Means Versions) Algorithms

- (1) It is obvious that the classification rate is quite inferior as compared to the labelled clustering algorithms

TABLE 4: Observations made on the classification capability of OPNN and APNN for moderate database—with clustering algorithm.

Preprocessing Technique	Misclassifications in OPNN with LVQ1	Misclassifications in OPNN with LVQ2	Misclassifications in OPNN with LVQ3
$\Phi$ - $q_{\max}$ - $n$ ( $30^\circ$ )	4 types—3 numbers (AC2, AC9, EC1) 5 types—6 numbers (EC1, EC3, EC7SD7, EC8SD8, EC2, EC4)	4 types—2 numbers (EC2, EC1) 5 types—3 numbers (EC1, EC6SD6, EC8SD8)	4 types—2 numbers (EC2, EC1) 5 types—4 numbers (EC2, EC1, EC7SD7, EC8SD8)
$\Phi$ - $q_{\min}$ - $n$ ( $30^\circ$ )	4 types—6 numbers (AC7, AC5, AC2, EC6AC7, EC7AC8, EC8AC9) 5 types—6 numbers (EC6C7, EC7AC8, EC7SD7, EC8SD8, AC2, EC5AC2)	4 types—3 numbers (EC1, EC6SD6, EC8SD8) 5 types—8 numbers (AC2, EC5, SD7, EC6AC7, EC7AC8, EC6SD6, EC8SD8)	4 types—5 numbers (AC2, EC5, EC6AC7, EC7AC8, EC8AC9) 5 types—8 numbers (AC2, EC5, SD7, EC6AC7, EC7AC8, EC8AC9, EC7SD7, EC8SD8)
$\Phi$ - $q_{\max}$ - $n$ ( $10^\circ$ )	4 types—3 numbers (EC2, EC8SD8, EC1) 5 types—7 numbers (EC2, EC1AC2, EC5AC2, EC1, EC3, EC7SD7, EC8SD8)	4 types—1 number (EC1) 5 types—4 numbers (EC1AC2, EC6AC2, EC6SD6, EC8SD8)	4 types—3 numbers (EC2, EC5AC2, EC1) 5 types—5 numbers (EC1AC2, EC6AC2, EC7SD7, EC8SD8)
$\Phi$ - $q_{\min}$ - $n$ ( $10^\circ$ )	4 types—2 numbers (AC2, AC7) 5 types—4 numbers (AC2, AC7, EC7SD7, EC8SD8)	4 types—2 numbers (EC5SD5, EC6AC2) 5 types—4 numbers (EC1AC2, EC6AC2, EC6SD6, EC8SD8)	4 types—2 numbers (AC1, AC2) 5 types—5 numbers (EC1AC2, EC6AC2, EC7SD7, EC8SD8)

during the training phase since it is an established fact that the selection of the initial seed (which is a random selection) would be vital for appropriate learning. However, the ability of such algorithms to be able to provide class separable boundaries, proposes an attractive alternative for input data validation and in addition to providing plausible solutions for identifying unknown categories. It is relevant to note that since the scope of the research is on assessing the capability of the clustering algorithms in providing solutions to handle large training datasets, only the more popular and traditional types of clustering algorithms have been implemented to ascertain this fact. However, a wide gamut of other improved versions of  $K$ -means algorithms (improved  $K$ -means, Greedy  $K$ -means etc.) may be attempted for better classification capabilities. Table 7 summarizes the important observations during the analysis of the classification capability of the unlabelled clustering algorithms.

- (2) It is also substantiated that improved classification rate is noticed for pre-processing scheme that utilizes the inequality relationship based on the measures pertaining to the types of mean values. This aspect was also noticed in the case of labelled clustering algorithms.

#### 8.6. Case Study 6: Capability of the Novel Hypergraph-PNN (HGPNN) in Classifying Multisource PD Patterns

- (1) It is also observed that the novel HG-PNN classifier serves as a significantly good center selection

algorithm though only a modest set of centers were obtained for classification. Table 8 clearly elucidates this aspect of utilizing the novel method of HG as a clustering algorithm in PD pattern recognition.

- (2) The best classification during the studies was obtained for values of the smoothing parameter within the range of 15–30. This characteristic delineates the fact that the separation of class boundaries is much wider than the previous studies carried out by the authors [14] of this research on similar set of testing of multi-source PD thus providing an index of good set of centers that represent the class of PD.
- (3) It is also obvious from Tables 8 and 9 that though the HGPNN performed outstandingly, the number of centers created by the HG algorithm is substantial as compared to the density estimation-based clustering/center selection algorithm for studies carried out by the authors of this research earlier [13]. This aspect could be ascribed to the utility of one of the properties of HG, namely, the Helly property. Yet, since the focus of the research is mainly to ascertain the capability of the HG algorithm in being adaptable as a center selection technique, other more salient properties of HG such as transversal, conformal, and mosaic have not been attempted.
- (4) In the case of measures based on  $\varphi$ - $q_{\max}$ - $n$  ( $30^\circ$ ), the number of centers obtained were much higher than the number of centers achieved by HG algorithm for measures based on  $\varphi$ - $q_{\min}$ - $n$  ( $10^\circ$ ). It is of significance to note that the classification capability of measures with  $30^\circ$  window was better than that of classification

TABLE 5: Comparison on the role of variance parameter in classifying multiple PD sources.

Input feature vector		APNN (without clustering)	OPNN with LVQ1	OPNN with LVQ2	OPNN with LVQ3
$\Phi$ - $q_{\max}$ - $n$ ( $10^\circ$ )	4 input types	$\sigma'_1 = 0.097$	$\sigma'_1 = 0.018$	$\sigma'_1 = 0.019$	$\sigma'_1 = 0.033$
		$\sigma'_2 = 0.195$	$\sigma'_2 = 0.025$	$\sigma'_2 = 0.021$	$\sigma'_2 = 0.040$
		$\sigma_3 = 0.058$	$\sigma_3 = 0.070$	$\sigma_3 = 0.073$	$\sigma_3 = 0.062$
		$\sigma_4 = 0.079$	$\sigma_4 = 0.020$	$\sigma_4 = 0.025$	$\sigma_4 = 0.016$
	5 input types	$\sigma'_1 = 0.171$	$\sigma'_1 = 0.024$	$\sigma'_1 = 0.037$	$\sigma'_1 = 0.031$
		$\sigma'_2 = 0.241$	$\sigma'_2 = 0.025$	$\sigma'_2 = 0.008$	$\sigma'_2 = 0.008$
		$\sigma_3 = 0.067$	$\sigma_3 = 0.079$	$\sigma_3 = 0.038$	$\sigma_3 = 0.040$
		$\sigma_4 = 0.078$	$\sigma_4 = 0.032$	$\sigma_4 = 0.028$	$\sigma_4 = 0.029$
		$\sigma_5 = 0.173$	$\sigma_5 = 0.016$	$\sigma_5 = 0.004$	$\sigma_5 = 0.003$
$\Phi$ - $q_{\min}$ - $n$ ( $10^\circ$ )	4 input types	$\sigma'_1 = 0.138$	$\sigma'_1 = 0.044$	$\sigma'_1 = 0.038$	$\sigma'_1 = 0.054$
		$\sigma'_2 = 0.206$	$\sigma'_2 = 0.015$	$\sigma'_2 = 0.021$	$\sigma'_2 = 0.009$
		$\sigma_3 = 0.071$	$\sigma_3 = 0.035$	$\sigma_3 = 0.043$	$\sigma_3 = 0.028$
		$\sigma_4 = 0.075$	$\sigma_4 = 0.011$	$\sigma_4 = 0.019$	$\sigma_4 = 0.007$
	5 input types	$\sigma'_1 = 0.172$	$\sigma'_1 = 0.032$	$\sigma'_1 = 0.037$	$\sigma'_1 = 0.032$
		$\sigma'_2 = 0.258$	$\sigma'_2 = 0.006$	$\sigma'_2 = 0.007$	$\sigma'_2 = 0.007$
		$\sigma_3 = 0.009$	$\sigma_3 = 0.037$	$\sigma_3 = 0.0307$	$\sigma_3 = 0.039$
		$\sigma_4 = 0.094$	$\sigma_4 = 0.027$	$\sigma_4 = 0.0277$	$\sigma_4 = 0.028$
		$\sigma_5 = 0.141$	$\sigma_5 = 0.003$	$\sigma_5 = 0.004$	$\sigma_5 = 0.003$

TABLE 6: Comparison of classification capability of OPNN and APNN with LVQ versions' clustering algorithms.

Input feature vector	Classification capability																	
	LVQ1 clustering				LVQ2 clustering						LVQ3 clustering							
	OPNN		APNN		OPNN		APNN				OPNN		APNN					
	Iter.	(%)	Iter.	(%)	Iter.	$\epsilon$	(%)	Iter.	$\epsilon$	(%)	Iter.	$\epsilon$	$\eta$	(%)	Iter.	$\epsilon$	$\eta$	(%)
$\Phi$ - $q_{\max}$ - $n$ ( $30^\circ$ )	5000	89	5000	91	500	.5	95	1000	.5	94	1000	.6	.3	93	1000	.7	.3	93
$\Phi$ - $q_{\max}$ - $n$ ( $10^\circ$ )	5000	87	5000	88	1000	.6	90	1000	.7	93	1000	.8	.2	89	1000	.7	.2	89
$\Phi$ - $q_{\min}$ - $n$ ( $30^\circ$ )	1000	91	1000	93	1000	.6	94	1000	.8	93	5000	.7	.3	92	1000	.6	.3	93
$\Phi$ - $q_{\min}$ - $n$ ( $10^\circ$ )	1000	88	1000	89	1000	.7	91	1000	.7	93	1000	.7	.3	91	1000	.8	.3	92
Measure of types of mean ( $30^\circ$ )	1000	91	1000	93	1000	.8	95	1000	.8	96	1000	.7	.3	92	3000	.7	.3	93
Measure of types of mean ( $10^\circ$ )	1000	92	1000	93	1000	.8	95	1000	.8	96	1000	.7	.3	92	1000	.8	.3	94

TABLE 7: Comparison of classification capability of OPNN and APNN with versions of  $K$ -means clustering algorithms.

Serial number	Input feature vector	Classification capability											
		PNN versions (without clustering)		OPNN (with standard $K$ -means clustering)				APNN (with the Forgy $K$ -means clustering)					
				OPNN		APNN		OPNN		APNN			
		OPNN	APNN	Iter.	(%)	Iter.	(%)	Iter.	(%)	Iter.	(%)	Iter.	(%)
(1)	$\Phi$ - $q_{\max}$ - $n$ ( $30^\circ$ )	93.6%	93%	1000	81	1000	83	2000	80	5000	81		
(2)	$\Phi$ - $q_{\max}$ - $n$ ( $10^\circ$ )	94%	93.3%	1000	84	1000	84	2000	81	5000	82		
(3)	$\Phi$ - $q_{\min}$ - $n$ ( $30^\circ$ )	89%	91%	5000	80	1000	81	5000	80	5000	81		
(4)	$\Phi$ - $q_{\min}$ - $n$ ( $10^\circ$ )	91%	92%	5000	81	1000	82	5000	81	5000	81		
(5)	Measure of types of mean ( $30^\circ$ )	94	94	5000	84	1000	85	5000	82	5000	82		
(6)	Measure of types of mean ( $10^\circ$ )	94	95	5000	85	1000	86	5000	83	5000	83		

TABLE 8: Optimal centers obtained from HG algorithm for PD pattern classification.

Preprocessing technique	Number of optimal centers from HG algorithm			
	Electrode bounded cavity	Air corona	Oil corona	Multiple sources
$\Phi$ - $q_{\max}$ - $n$ ( $30^\circ$ )	7.3 kV-26	14 kV-17	21 kV-18	7.3 kV-11
	9.1 kV-13	21 kV-15	29.1 kV-17	9.1 kV-12
	9.6 kV-7	23 kV-8	32 kV-14	10 kV-17
$\Phi$ - $q_{\max}$ - $n$ ( $10^\circ$ )	7.3 kV-6	14 kV-8	21 kV-15	7.3 kV-8
	9.1 kV-16	21 kV-12	29.1 kV-16	9.1 kV-15
	9.6 kV-10	23 kV-10	32 kV-14	10 kV-17
$\Phi$ - $q_{\min}$ - $n$ ( $10^\circ$ )	7.3 kV-12	14 kV-16	21 kV-12	7.3 kV-18
	9.1 kV-18	21 kV-18	29.1 kV-13	9.1 kV-21
	9.6 kV-8	23 kV-17	32 kV-15	10 kV-17
AM-GM-HM-RM ( $10^\circ$ )	7.3 kV-9	14 kV-15	21 kV-17	7.3 kV-17
	9.1 kV-26	21 kV-18	29.1 kV-17	9.1 kV-16
	9.6 kV-9	23 kV-17	32 kV-19	10 kV-15

TABLE 9: Classification capability of HGPNN for multiple source PD patterns.

Preprocessing scheme	Phase window	No. of tuples	Training patterns	Classification capability (%)
Measures based on maximum	$\Phi$ - $q_{\max}$ - $n$ ( $30^\circ$ )	36	175	97
Measures based on maximum value	$\Phi$ - $q_{\max}$ - $n$ ( $10^\circ$ )	36	174	96.67
Measures based on minimum	$\Phi$ - $q_{\min}$ - $n$ ( $10^\circ$ )	36	188	93.6
Measures based on mean	AM-GM-HM-RM ( $10^\circ$ )	36	186	90.5

based on  $10^\circ$ . However, it is obvious that this has been achieved at the cost of higher number of centers as observed in Table 8.

- (5) Tables 8 and 9 clearly enunciate the fact that the number of centers that essentially describe the source of PD is dependent on the dimensionality of the HG centers. It is evident that the classification capability is enhanced with number of representative centers while a slightly inferior classification rate is obtained for a larger dimensionality (tuple) though with substantially larger number of centers. Though “curse of dimensionality” is a vital aspect in designing computationally effective clustering algorithms, the nature of centers obtained provides a much broader value of the smoothing parameter, thus circumventing the stated aspect previously discussed.

*8.7. Comparison of Classification Capacity of HGPNN with Feedforward Backpropagation (FFBPA) Neural Network.* Preliminary studies carried out by the authors of this research earlier [33] clearly indicate limitations pertaining to long training epoch (in several cases prohibitively large training time in the range of 8–10 hours) for convergence during the iterative procedure even in the case of small dataset training. Since large dataset training and testing is taken up for studies in this research, it is obvious that the training phase would necessitate more robust training strategies for better computational cost. These observations also clearly

indicate the limitations during the training phase of the FFBPA network as discussed in [58] where the research findings of Specht and Shapiro deliberate this aspect. This issue becomes even more significant in the context of training and testing large dataset, online, complex real time PD signature analysis.

*8.8. Comparison of Classification Capacity of HGPNN with Wavelet Transform-PNN Classifier.* In this context, for the purpose of comparison, studies based on discrete wavelet transformation (DWT) have also been taken up in this work since recent studies by researchers have indicated the merits of utilizing this technique in discriminating overlapped PD signatures most prevalent during practical measurements on-site. The Daubechies wavelet has been utilized in this work as it has been observed that this family of wavelets has desirable properties that usually match the requirements pertaining to PD pattern classification such as data compression and compactness, orthogonality, and asymmetry for analysis for fast varying pulses. Since, a few classical studies based on wavelet transformation in PD analysis [20] also provide substantial guidelines in the appropriate selection of the order and level of the selected wavelet, it is found relevant to use higher-order and lower-level (scale) wavelet representation for pattern recognition tasks. Hence, in this study the Daubechies wavelet with order 7 and level 3 was taken up for obtaining the approximate and detailed coefficients. Based on the coefficients obtained, postprocessing and further studies have been carried out

TABLE 10: Capability of wavelet transform-PNN in classifying multiple source PD signatures.

Feature vector	Number of tuples	Total number of windows for statistically extracted features	Total number of PD signatures	Classification capability (%)			
				OPNN without clustering	APNN without clustering	LVQ2 clustering OPNN	APNN
The Daubechies coefficients (order 7 and level 3)	192	16	480	92	93.1	94.2	94.7
	264	9	480	90.2	91.3	92.3	93.1

utilizing statistical measures (range, standard deviation, mean, skewness, and kurtosis) for a phase window of 30° and 10°. Table 10 summarizes the analysis carried out utilizing wavelet transform.

It is obvious from Table 10 that the number of feature extraction bins (during the extraction of the wavelet coefficients based on statistically processed measures) plays a vital role in the capability of classification of the WT-PNN. It is pertinent to observe that with increased dimensionality of the extracted features, the classification capability is not enhanced, in fact, detrimental to classification. This aspect clearly exemplifies the need for appropriate center selection strategies (such as HG-based clustering).

Further it is evident from the detailed analysis and from case study shown in Table 10 that good classification capability of the wavelet PNN is obtained for considerably larger number of tuples of extracted features as compared to considerably lesser-dimensioned features obtained from simple statistical measures based on HG methodology. Thus much more parsimonious sets of centers are obtained with more compact feature representatives with the HG-based center selection and clustering technique though with slightly inferior classification capability. However, it would be worth mentioning in this context that this limitation may be attributed to the utility and exploitation of only one of the preliminary property of HG, namely, Helly, while several other powerful salient properties of HG such as transversal, mosaic, and conformal have not been taken up in this research. Such properties are expected to provide enhanced results.

## 9. Conclusions

The role played by both partition and graph theory-based clustering algorithms in discriminating multi-source PD patterns utilizing the two basic variants of PNN are summarized as follows.

- (1) During the training phase-labelled versions of LVQ clustering augurs well as a good learning scheme and are able to handle ill-conditioned dataset and overlapped multiple PD sources considerably well. It is also evident that this method may be appropriate during offline studies wherein under controlled testing conditions, appropriate training of prototype vectors pertaining to a particular class would ensure a compact and reasonable codebook vector for further classification by PNNs.

- (2) The unlabelled clustering algorithm offers fresh insight into possible schemes for cluster validation which may consequently present a likely methodology for recognition of unknown class of PD sources during real time studies. Though this scheme may appear to be more associated with its counterpart (weak learning strategy), it is essential to note that since PD source discrimination is fundamental for successful insulation diagnosis it may be reasonable that the sources of PD signatures are classified from the viewpoint of strong learning strategy. The authors of this research are engaged in attempting a cluster validation-based scheme which is ongoing presently.
- (3) It is evident from the studies that HG-based center selection/clustering algorithm provides an exciting and a viable option for obtaining reasonably parsimonious set centers that describe the class of PD. Though the properties of the HG algorithm was utilized only to cluster and classify the PD patterns in this research, this scheme provides an exciting opportunity to correlate the relationship/association of PD pulses in terms of geometric aspects also. This research aspect is presently ongoing. Since much larger sets of representative centers are observed during this study, more appropriate properties of HG such as transversal, conformal, and mosaic can be attempted to further validate the approach.

## Acknowledgments

This research was supported by the Research and Modernization Fund (RMF) Grant, Project no. 6, constituted by the SASTRA University. The first author is extremely grateful to Professor Sethuraman, Vice-Chancellor, Dr. S. Vaidhyasubramaniam, Dean-Planning and Development, and Dr. S. Swaminathan, Dean-Sponsored Research and Director-CeNTAB, SASTRA University for awarding the grant and for the unstinted support and motivation extended to him during the course of the project. The authors reminisce Dr. P. S. Srinivasan, formerly Dean/SEEE, SASTRA University for many useful discussions and suggestions.

## References

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

- [2] R. Bartnikas, "Partial discharges their—mechanism, detection and measurement," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 9, no. 5, pp. 763–808, 2002.
- [3] S. Senthil Kumar, M. N. Narayanachar, and R. S. Nema, "Pulse sequence studies on PD data," in *Proceedings of the 11th International Symposium on High Voltage Engineering*, pp. 5.25.S1–5.25.S.7, UK, 1999.
- [4] E. Gulski and A. Krivda, "Neural networks as a tool for recognition of partial discharges," *IEEE transactions on electrical insulation*, vol. 28, no. 6, pp. 984–1001, 1993.
- [5] A. A. Mazroua, R. Bartnikas, and M. M. A. Salama, "Discrimination between PD pulse shapes using different neural network paradigms," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 1, no. 6, pp. 1119–1130, 1994.
- [6] L. Satish and W. S. Zaengl, "Artificial neural networks for recognition of 3-D partial discharge patterns," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 1, no. 2, pp. 265–275, 1994.
- [7] M. M. A. Salama and R. Bartnikas, "Determination of neural-network topology for partial discharge pulse pattern recognition," *IEEE Transactions on Neural Networks*, vol. 13, no. 2, pp. 446–456, 2002.
- [8] B. Karthikeyan and S. Gopal, "A novel complex probabilistic neural network system for classification of partial discharge patterns," in *Proceedings of the 14th International Symposium on High Voltage Engineering*, pp. 25–29, Beijing, China, August 2005.
- [9] B. Karthikeyan, S. Gopal, and S. Venkatesh, "Probabilistic neural network and its adaptive version—a stochastic approach to pd pattern classification task," *International Journal of Information Acquisition*, vol. 2, no. 4, pp. 1–12, 2005.
- [10] B. Karthikeyan, S. Gopal, and S. Venkatesh, "A heuristic complex probabilistic neural network system for partial discharge pattern classification," *Journal of the Indian Institute of Science*, vol. 85, no. 5, pp. 279–294, 2005.
- [11] B. Karthikeyan, S. Gopal, and S. Venkatesh, "Partial discharge pattern classification using composite versions of probabilistic neural network inference engine," *Expert Systems with Applications*, vol. 34, no. 3, pp. 1938–1947, 2008.
- [12] T. K. Abdel-Galil, R. M. Sharkawy, M. M. A. Salama, and R. Bartnikas, "Partial discharge pattern classification using the fuzzy decision tree approach," *IEEE Transactions on Instrumentation and Measurement*, vol. 54, no. 6, pp. 2258–2263, 2005.
- [13] A. Contin, A. Cavallini, G. C. Montanari, G. Pasini, and F. Puletti, "Digital detection and fuzzy classification of partial discharge signals," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 9, no. 3, pp. 335–348, 2002.
- [14] L. Satish and W. S. Zaengl, "Can fractal features be used for recognizing 3-D partial discharge patterns?" *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 3, pp. 352–359, 1995.
- [15] E. M. Lalitha and L. Satish, "Fractal image compression for classification of PD sources," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 5, no. 4, pp. 550–557, 1998.
- [16] T. K. Abdel-Galil, Y. G. Hegazy, M. M. A. Salama, and R. Bartnikas, "Partial discharge pulse pattern recognition using hidden Markov models," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 11, no. 4, pp. 715–723, 2004.
- [17] L. Satish and B. I. Gururaj, "Use of hidden Markov models for partial discharge pattern classification," *IEEE transactions on electrical insulation*, vol. 28, no. 2, pp. 172–182, 1993.
- [18] S. Venkatesh, S. Gopal, and K. Kannan, "A novel hybrid continuous density hidden markov model—probabilistic neural network for multiple source partial discharge pattern recognition," in *Proceedings of the 17th International Symposium on High Voltage Engineering (ISH'11)*, p. 402, F-039, Hannover, Germany, August 2011.
- [19] E. M. Lalitha and L. Satish, "Wavelet analysis for classification of multi-source PD patterns," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 7, no. 1, pp. 40–47, 2000.
- [20] X. Ma, C. Zhou, and I. J. Kemp, "Interpretation of wavelet analysis and its application in partial discharge detection," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 9, no. 3, pp. 446–457, 2002.
- [21] J. H. Lee, T. Okamoto, and C. W. Yi, "Classification of PD patterns from multiple defects," in *Proceeding of the 6th International Conference on Properties and Applications of Dielectric Materials*, pp. 463–465, June 2000.
- [22] R. J. Van Brunt, "Stochastic properties of partial discharge phenomenon," *IEEE Transactions on Electrical Insulation*, vol. 26, no. 5, pp. 902–948, 1991.
- [23] R. J. Van Brunt and E. W. Cernyar, "Importance of unraveling memory propagation effects in interpreting data on partial discharge statistics," *IEEE Transactions on Electrical Insulation*, vol. 28, no. 6, pp. 905–916, 1993.
- [24] M. G. Danikas and A. D. Karlis, "On the use of neural networks in recognizing sources of partial discharge in electrical machine insulation: a short review," *International Review of Electrical Engineering*, vol. 1, no. 2, pp. 277–285, 2006.
- [25] E. Gulski, "Digital analysis of partial discharges," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 5, pp. 822–837, 1995.
- [26] A. Krivda, "Automated recognition of partial discharges," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 5, pp. 796–821, 1995.
- [27] R. E. James and B. T. Phung, "Development of computer-based measurements and their application to PD pattern analysis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 5, pp. 838–856, 1995.
- [28] R. Candela, G. Mirelli, and R. Schifani, "PD recognition by means of statistical and fractal parameters and a neural network," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 7, no. 1, pp. 87–94, 2000.
- [29] A. A. Mazroua, M. M. A. Salama, and R. Bartnikas, "PD Pattern recognition with neural networks using the multilayer perceptron technique," *IEEE Transactions on Electrical Insulation*, vol. 28, no. 6, pp. 1082–1089, 1993.
- [30] N. B. Bish, P. A. Howson, R. J. Howlett, T. J. Fawcett, and D. A. Hilder, "Combined intelligent PD analysis of high voltage dielectric condition evaluation," in *Proceedings of 11th International Symposium on High Voltage Engineering (ISH'99)*, London, UK, 1999.
- [31] M. Hoof, B. Freisleben, and R. Patsch, "PD source identification with novel discharge parameters using counterpropagation neural networks," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 4, no. 1, pp. 17–32, 1997.
- [32] B. Karthikeyan, S. Gopal, and S. Venkatesh, "ART 2-an unsupervised neural network for PD pattern recognition and classification," *Expert Systems with Applications*, vol. 31, no. 2, pp. 345–350, 2006.
- [33] B. Karthikeyan, S. Gopal, S. Venkatesh, and S. Saravanan, "PNN and its adaptive version—an ingenious approach to PD

- pattern classification compared with BPA network,” *Journal of Electrical Engineering*, vol. 57, no. 3, pp. 138–145, 2006.
- [34] S. Venkatesh, S. Gopal, P. S. S. Srinivasan, and B. Karthikeyan, “Identification of multiple source partial discharge patterns using bayesian classifier based neural networks—a comparison of supervised and unsupervised learning techniques,” in *Proceedings of 15th International Symposium of High Voltage Engineering, Slovenia (ISH’07)*, no. T5-407, p. 198, 2007.
- [35] D. F. Specht, “Probabilistic neural networks for classification, mapping or associative memory,” in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 1, no. 1, pp. 525–532, 1998.
- [36] D. F. Specht, “Probabilistic neural networks and the polynomial Adaline as complementary techniques for classification,” *IEEE Transactions on Neural Networks*, vol. 1, no. 1, pp. 111–121, 1990.
- [37] C. H. Chen, *Fuzzy Logic & Neural Network Handbook*, McGraw-Hill, New York, NY, USA, 1st edition, 1996.
- [38] T. Masters, *Advanced Algorithms for Neural Networks: A C++ Sourcebook*, John Wiley & Sons, New York, NY, USA, 1995.
- [39] S. Venkatesh and S. Gopal, “Robust heteroscedastic probabilistic neural network for multiple source partial discharge pattern recognition—significance of outliers on classification capability,” *Expert Systems with Applications*, vol. 38, no. 9, pp. 11501–11514, 2011.
- [40] S. Venkatesh and S. Gopal, “Orthogonal least square center selection technique—a robust scheme for multiple source Partial Discharge pattern recognition using Radial Basis Probabilistic Neural Network,” *Expert Systems with Applications*, vol. 38, no. 7, pp. 8978–8989, 2011.
- [41] D. F. Specht and H. Romsdahl, “Experience with adaptive probabilistic neural networks and adaptive general regression neural networks,” in *Proceedings of the IEEE International Conference on Neural Networks*, pp. 1203–1208, June 1994.
- [42] N. K. Bose and P. Liang, *Neural Network Fundamentals With Graphs, Algorithms and Applications*, McGraw-Hill, Hightstown, NJ, USA, 1996.
- [43] V. Faber, “Clustering and the continuous  $K$ -means algorithm,” *Journal of Los Alamos Science*, vol. 22, pp. 38–144, 1994.
- [44] R. Xu and D. Wunsch, “Survey of clustering algorithms,” *IEEE Transactions on Neural Networks*, vol. 16, no. 3, pp. 645–678, 2005.
- [45] Z. Tian, T. Hwang, and R. Kuang, “A hypergraph-based learning algorithm for classifying gene expression and arrayCGH data with prior knowledge,” *Bioinformatics*, vol. 25, no. 21, pp. 2831–2838, 2009.
- [46] L. Fausett, *Fundamentals of Neural Networks—Architectures, Algorithms and Applications*, Pearson Education, 2004.
- [47] P. Burrascano, “Learning vector quantization for the probabilistic neural network,” *IEEE Transactions on Neural Networks*, vol. 2, no. 4, pp. 458–461, 1991.
- [48] H. S. Cho, “Opto-mechatronic handbook: techniques and applications,” in *Pattern Recognition*, chapter 9, CRC Press, Fla, USA, 2003.
- [49] A. Bretto, “Introduction to hypergraph theory and its use in engineering and image processing,” in *Advances in Imaging and Electron Physics*, vol. 131, Chapter 1, Elsevier Academic Press, 2004.
- [50] D. Zhou, J. Huang, and B. Scholkopf, “Learning with hypergraphs: clustering, classification and embedding,” in *Proceedings of the Advances in Neural Information Processing Systems (NIPS’06)*, pp. 1601–1608, MIT Press, Cambridge, Mass, USA, 2006.
- [51] J. S. Cherng and M. J. Lo, “A hypergraph based clustering algorithm for spatial data sets,” in *Proceedings of the 1st IEEE International Conference on Data Mining (ICDM’01)*, pp. 83–90, San Jose, Calif, USA, December 2001.
- [52] IEC, 60270, High Voltage Test Techniques-Partial Discharge Measurements, 2000.
- [53] CIGRE Working Group, “Recognition of discharges,” Report 21.03 Electra No. 11, 1969.
- [54] T ASTM, D, 61847, Tentative methods of conditioning plastics and electrical insulating materials for testing, 2003.
- [55] W.-S. Lim and M. V. V. Rao, “A new method of reducing network complexity in probabilistic neural network for target identification,” *IEICE Electronics Express*, vol. 1, no. 17, pp. 534–539, 2004.
- [56] R. A. Fischer, “The use of multiple measurements in taxonomic problems,” *Annual Eugenics*, vol. 7, Part II, pp. 179–188, 1936.
- [57] Ke Lin Du, M. N. S. Swamy, and K. L. Du, *Neural Networks in a Soft Computing Framework*, chapter 6, Springer, London, 1st edition, 2006.
- [58] D. F. Specht and P. D. Shapiro, “Generalization accuracy of probabilistic neural networks compared with back-propagation networks,” in *Proceedings of IEEE International Conference on Neural Networks*, pp. 887–892, Seattle, Wash, USA, July 1991.



# Hindawi

Submit your manuscripts at  
<http://www.hindawi.com>

