

# ROLE OF PARTITIONING BASED CLUSTERING ALGORITHMS IN CLASSIFYING MULTI-SOURCE PARTIAL DISCHARGE PATTERNS USING PROBABILISTIC NEURAL NETWORK AND ITS ADAPTIVE VERSION – A REVIEW

<sup>1</sup>S. VENKATESH, <sup>2</sup>S. GOPAL, <sup>3</sup>S. JAYALALITHA

<sup>1</sup>Asstt Prof., Department of EEE, SEEE, SASTRA University, Thanjavur, INDIA-613401

<sup>2</sup>Executive Director, W.S. Test Systems Pvt. Limited, Bangalore, INDIA-562157

<sup>3</sup>Assoc. Dean, Department of EIE, SEEE, SASTRA University, Thanjavur, INDIA-613401

E-mail: [venkatsri73in@gmail.com](mailto:venkatsri73in@gmail.com), [drsgopal@wstestsystems.com](mailto:drsgopal@wstestsystems.com), [sj\\_instru@eie.sastra.edu](mailto:sj_instru@eie.sastra.edu)

## ABSTRACT

Notwithstanding developments in design, processing and quality control techniques, flaws in insulation of electrical power apparatus such as blow-holes, cavities, surface imperfections, etc are unavoidable. Manufacturers and utilities employ a variety of prognostic tools to assess the condition of insulation amongst which Partial Discharge (PD) measurement has become an indispensable non-intrusive technique. Since discrimination of PD is a fundamental yet an indispensable prerequisite for diagnosis and since multiple PD sources are frequently encountered in practice, such studies have grabbed the attention of researchers. In view of the fact that PD is inherently a complex non-Markovian process and a majority of modern PD measurement systems acquire data for a substantial duration, the database is large leading to a plethora of complications during classification such as ill-conditioning, over-fitting etc. Though an array of intelligent and scientific computational techniques such as Neural Networks (NN), Hidden Markov Models, Wavelet Transformation etc have been utilized for single and overlapped PD sources with a fair degree of success, the inability to develop a comprehensive and validated scheme for identification continues to be elusive. In this research, exhaustive studies have been carried out on benchmark models that replicate single and multiple PD sources with a focus on three major aspects during classification which utilizes PNN versions that employ a variety of partition based clustering techniques namely algorithms based on distance and similarity, square error and density function. The first aspect focuses on ascertaining the ability of non-clustering based PNN versions in handling large ill-conditioned datasets during training in comparison with the various partitioning based algorithms. The second is on the analysis of the ability of square-error, distance and density based clustering techniques in providing frugal sets of representative centers during training. The third facet is to ascertain the role played by the preprocessing techniques in dealing with issues related to the curse of dimensionality during classification. Further, exhaustive analysis is carried out to determine the role played by the free parameter (variance parameter) in distinguishing various classes of PD, 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 codebook vectors in classifying the patterns.

**Keywords:** *Neural Network (NN), Probabilistic Neural Network (PNN), Adaptive Probabilistic Neural Network (APNN), Learning Vector Quantization (LVQ), Expectation Maximization (EM)*

## 1. INTRODUCTION

Partial Discharge (PD) [1] is an electrical breakdown phenomenon restricted to the localized region of the insulating system of a power apparatus. PD, which may consequently lead to deterioration of the insulation system due to chemical degradation of a power apparatus, occurs

as internal discharges in cavities, voids, blow-holes, gaps at the interfaces etc., as external or gliding discharges on the surface imperfections, as corona discharges at sharp points and protrusions etc. Hence, it is of considerable practical relevance to researchers and operators handling utilities, to be able to distinguish the sources of PD, its geometry and location since such aspects are intimately

related to the condition monitoring and diagnosis of the insulation system of the equipments. Due to a surfeit of advances in digital hardware systems, improvements in the computational speed of processors and advancements in associated data processing and acquisition systems there has been a reorientation among researchers in carrying out PD measurement and analysis. More-so, in recent years, the trend has shifted to identification and discrimination of PD signatures due to multiple sources of PD [2-3] since these are often encountered during on-site measurements wherein distinguishing various sources of PD becomes progressively demanding.

A broad range of transform based techniques such as Artificial Neural Networks (ANN) [3-8], Fuzzy Logic Controller (FLC) [9-10], Fractal Features [11-12], Hidden Markov Models [13-15], Fast Fourier Transform (FFT), Wavelet Transform [16-17] etc., have been utilized by researchers to identify and determine defects in the insulation system. Despite the fact that efforts to classify single source PD and partially overlapped patterns [16] have resulted in a fair degree of success, intricacies in classifying fully overlapped PD patterns in real-time insulation systems, the non-Markovian characteristic of patterns [18-19], disparity in the PD patterns due to varying applied voltages in realistic systems etc., still continue to put-forth considerable impediments [20-21].

Detailed laboratory analysis has been taken up on benchmark models as stipulated in [22] which would in turn serve as a validation scheme in ascertaining the capability and reliability of the proposed clustering and NN algorithms in identifying the complex PD signatures. This research work focuses on three major aspects during the classification of multiple PD sources utilizing PNN versions that employ a variety of partition based clustering techniques namely the distance and similarity based algorithms, square error based algorithms and density function based algorithms. The principal thrust is on ascertaining the ability of non-clustering based versions of PNN in handling large ill-conditioned datasets during the training phase in comparison with the various partitioning based clustering algorithms. The second aspect of this study is the analysis of the ability of square-error, distance and density based clustering techniques in providing frugal sets of representative centers/ prototype vectors during the training phase. The third facet of this research is to ascertain the role played by the preprocessing and feature extraction techniques in addressing issues

pertaining to the curse of dimensionality in the classification task. Further, exhaustive analysis is carried out to determine the role played by the free parameter (variance parameter) in distinguishing the various classes of PD, 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. PD MEASUREMENT, FEATURE EXTRACTION AND NEURAL NETWORKS FOR PATTERN CLASSIFICATION- AN OVERVIEW

### 2.1 Pd Detection, Measurement And Acquisition

Though several attributes of PD such as magnitude, rise time, recurrence rate, phase-relationship, time interval between successive pulses, discharge inception and extinction voltage etc characterize its occurrence, the more appropriate features from the viewpoint of pattern recognition and also as represented in a majority of modern digital PD measurement and acquisition systems are the phase angle of occurrence of PD pulses ( $\phi$ ), magnitude of the apparent charge during discharge ( $q$ ) and the discharge rate ( $n$ ). PD detection and measurement is carried out essentially based on the nature of primary characteristics and its working phenomena and broadly classified into non-electrical and electrical methods. The non-electrical methods include optical, acoustic and chemical methods of which the acoustic method (Ultra High Frequency) is usually resorted to for on-site measurement of PD since electro-magnetic interference and external disturbances pose limitations for on-site application of the electrical detection method. The other classification which is based on the primary characteristics includes the number of pulses employed by the detector, the bandwidth and the method of the display system. If the number of input measures a voltage or a current at the measuring point of the test object it is called a single input or the straight detection method. Multi-input methods such as balanced bridge detection are resorted to when the signal to noise level is low.

PD detectors may be classified based on bandwidth into broadband and narrowband types, the distinction of which is not necessarily clearly defined and is usually related to the required pulse resolution. If the bandwidth of the detector is sufficiently wide to resolve the individual pulses as stipulated by regulatory standards, the detector is

classified as broadband. Commercial PD detectors are band-pass in nature. Broadband systems provide facility for detection and measurement of individual pulses and have bandwidths, which extend to several hundred Hertz. For finer details of pulses, ultra wideband detectors are utilized. However, such detectors suffer from the disadvantage of being affected by background noise and hence pose problems in on-site measurements. Narrow band systems have long time constants integrating the detected signals and hence individual pulses appear as average values. Narrowband detectors use bandwidths over a flexible range. They are better suited for on-site measurements since they are less affected by external disturbances. However, if the impulses are insufficiently separated, an indistinct signal appears on the screen of the oscilloscope and the individual charges cannot be discerned.

With regard to display processing systems, the basic meter display system (analog meter with the movement related to display quasi-apparent charge or digital panel meter) and later direct displays (giving 'signatures' of different types of PDs) are increasingly being replaced by computer driven displays [23]. The computer driven display method involves detecting PD and digitizing the magnitudes pertaining to the output waveform. The compiled measurement results in tabulated or graphical form of display on a computer monitor. With regard to recent developments in the area of computer driven display methods, digital processing using the concept of phase resolved observations of PD representation for evaluating different discharges has been used. An analyzer, which evaluates the phase position quantities, has been automated. This system consists of a PD detector circuit, a data acquisition unit and a computer. Measurement of PD is carried out usually by utilizing a commercial PD detector. The PD pulses are integrated such that the maximum value of the integrated signal is proportional to the apparent charge. The peak value of the memory with sample and hold is used to register continuously all PD peak values. The acquisition time of the peak value storage is adjusted to PD detector bandwidth. Because of the time dependent behavior of PD under AC voltage conditions, several quantities as function of phase angle can be used to describe PD phenomena. For this purpose the voltage cycle is divided into phase windows ( $0^\circ$  to  $360^\circ$ ) wherein the observation takes place over several cycles. In each window, the statistical distribution of individual PD events can be determined.

With the advent of high speed and large memory processing components and data acquisition systems, three dimensional representation of PD pulse patterns ( $\phi$ -  $q$ -  $n$ ) have made it possible to carry out PD analysis based on two major categories. They are: 1. phase-resolved and 2. time-resolved data. Phase resolved data provides information regarding the distribution of pulses in a constant time base (either on a sinusoidal or elliptical) and hence augurs well for pattern classification and defect discrimination. However, information regarding pulse shapes, providing relationship between the nature of defect and the shape of pulses cannot be deciphered using this method which is facilitated by the second method.

## 2.2 Feature Extraction

Over the past few decades a wide range of preprocessing and feature extraction schemes have been utilized for PD pattern classification. Such methods include statistical measures, pulse characteristic parameters, image processing parameters, signal processing based transformation parameters etc. Statistical measures comprise statistical moments (skewness and kurtosis) [24-27], measures based on dispersion (range, standard deviation, variance, quartile deviation etc), measures of central tendency (arithmetic mean, median, moving average etc), cross-correlation, discharge asymmetry etc. Pulse characteristic based techniques are usually used in time resolved PD studies which include parameters such as pulse rise time, decay time, pulse width, repetition rate, quadratic rate, peak discharge magnitude etc.

## 2.3 Neural Networks For Pd Pattern Recognition

Preliminary studies on PD data for pattern recognition which utilized the Multilayer Perceptron (MLP) based Feed Forward Neural Network (FFNN) with Back Propagation Algorithm (BPA) [28] for supervised training of the network was a substantial success. However, extensive studies indicated that the rudimentary schemes resulted in a substantial computational cost in terms of long training periods. Later, studies based on Radial Basis Function (RBF) neural networks [29] indicated better results with improved convergence rate in addition to better demarcation of the hyper boundaries of the feature vectors during the training phase. Yet further improvement in discrimination capability at reduced computational cost continues to elude researchers involved in PD pattern classification.

Incidentally, the tendency of researchers shifted towards utilizing unsupervised learning neural networks such as Learning Vector Quantization (LVQ), Self Organized Map (SOM), Adaptive Resonance Theory (ART) [30] etc which clearly indicated a reasonable level of success in classifying single source PD patterns. Yet complexities related due to the inherent non-Markovian nature of PD pulses compounded further by varying applied voltages in real-time operation, difficulties in classifying multiple sources of PD (partially and fully overlapped patterns) in practical insulation systems, inevitability of ill-conditioned large dataset which may become impracticable to handle during the training phase etc., necessitate better understanding towards implementing a simple yet effective NN as a tool for the classification task.

In this context it is evident from the studies taken up earlier by the authors of this research work in classifying small dataset PD using PNN and its adaptive version [31-36] which reported encouraging results and provided crucial indications on its feasibility in addressing issues discussed in the preceding sections that this scheme could provide plausible solution in the framework of being a simple yet an effective technique. This perception originates from the fact that the standard version of the PNN and its adaptive version is based on the Parzen window density estimation technique (obtaining optimal probability density estimate) with a Bayesian based decision classifier (conditional density estimate as a part of the decision layer) for obtaining the class separability with only the variance parameter to be tuned during the training phase. This makes the scheme a simple yet an effective strategy in obtaining rapid classification capability.

Hence, the focus of this research is on investigating the ability of PNN and its adaptive version in classifying multiple sources of PD without and with clustering algorithms at varying applied voltages. The performance of these algorithms in handling large and ill-conditioned datasets acquired from the digital PD measurement and acquisition system is also studied.

### 3. PROBABILISTIC NEURAL NETWORK AND ITS ADAPTIVE VERSION- A REVIEW

PNN is a network representation of ‘probability density estimation’ with a model description based on competitive learning with a ‘winner takes all’

approach [37-39]. The core concept is based on multivariate probability estimation. The original (OPNN) and the adaptive versions of PNN (APNN) do not have feedback paths. PNN is a classifier version, which combines the Bayesian strategy for decision-making with a non-parametric estimator (Parzen window) for obtaining the probability density function (PDF). The PNN as described in Fig. 1 is a four layer architecture comprising an input layer, two hidden layers (one each for exemplar and class) and an output layer.

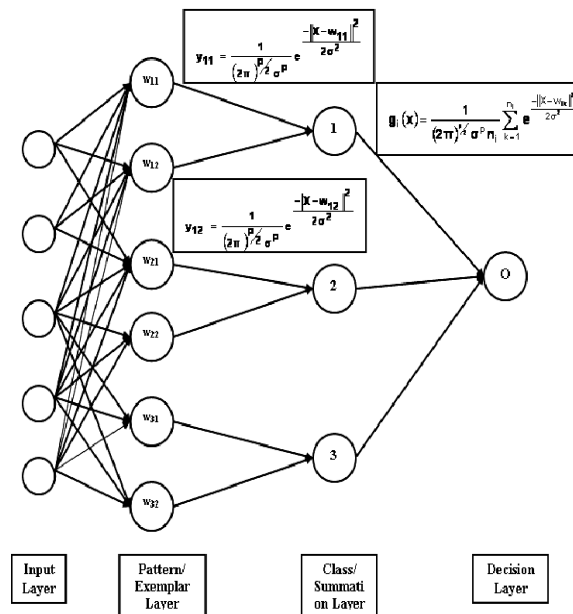


Fig. 1: Structure of Original Version of PNN

The pattern unit in Fig.1 may require normalization of input and exemplar vectors to unit length in many cases depending on the nature of data during the training phase. Various types of normalization procedures such as Euclidean, a special case of Minkowski (City Block), Mahalanobis etc may be resorted to of which the Euclidean and the city block norms are more prevalent. The pattern unit can be made independent of the requirement of unit normalization by adding the length of both vectors as inputs to the pattern unit.

#### 3.1 Adaptive PNN (APNN)

A noteworthy variation of the PNN called the Adaptive PNN (APNN) [40-41] exploits an effective methodology in utilizing the shaping parameter ‘σ’ (variance parameter) or the smoothing parameter. While the original PNN

utilizes a single value of smoothing parameter for all the classes for decision making, the adaptive PNN makes use of different values of  $\sigma$  for each class based on the calculation of an Euclidean distance and then computing the average distance ( $d_{ave}$ ) based on  $\sigma = g \cdot d_{ave}$  where 'g' is a constant which is incidentally the free parameter that can be tuned for adapting the variance in the dataset. Another major feature of this network is in implementation of a simplified formula of probability density function (PDF) which in turn obviates the need to normalize the input data resulting in considerable reduction in computation.

#### 4. CLUSTERING AND CENTER SELECTION ALGORITHMS: AN OVERVIEW

One of the quintessential requirements in dealing with datasets is to group them into a set of categories called clusters. In cluster analysis, such group of data is segregated into a number of homogeneous subgroups on the basis of an often subjectively chosen measure of similarity, such that the similarity between objects within a subgroup is more than the similarity between objects belonging to different subgroups. Hence, clustering involves dividing a set of data points into non-overlapping groups or clusters of points wherein the points in the cluster are 'more similar' to one another than to points in other clusters [42]. The term 'more similar' when applied to clustered points, usually means closer by measure of proximity. When a dataset is clustered, every point is assigned to some cluster and every cluster can be characterized by a single reference point, usually an average of the points in the cluster. Any particular division of all points in the dataset cluster is called partitioning. Data reduction is accomplished by replacing the coordinates of each point in a cluster with those of that cluster's reference point. The value of a particular clustering method will depend 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.

Clustering techniques are broadly classified as hierarchical and partition based clustering [43]. Hierarchical clustering technique groups the data with a sequence of partitions, from single cluster to a cluster including all clusters or vice versa, while partition clustering directly divides data objects into some pre-specified number of clusters without the hierarchical structure. Some of the major partition

based clustering techniques based on square error, density estimate include Vector Quantization, K-Means, Expectation Maximization (EM) with Maximum Likelihood (ML) etc. One of the important factors in partition clustering is the criterion function. 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. The sum of squared error function is one of the most widely used criteria. 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' 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.

From a probabilistic viewpoint, data objects are assumed to be generated according to several probability distributions. Data points in different clusters are generated by different probability distributions. They can be derived from different types of density functions (e.g., multivariate Gaussian). If the distributions are known, finding the clusters of a given data set is equivalent to estimating the parameters of several models. Maximum likelihood (ML) estimation is an important statistical technique for parameter estimation and it considers the best estimate as the one that maximizes the probability of generating all the observations, which in turn provides the joint density function.

#### 4.1 Learning Vector Quantization Versions Of Clustering Algorithms

Learning Vector Quantization (LVQ) developed by Kohonen [44] is fundamentally a supervised learning technique for pattern classification in which each output neuron represents a particular class or category. The weight vector for an output neuron is generally referred to as a 'codebook' vector of the class. During the training stage, the output units are positioned by adjusting the weight vector to approximate the decision hyper surface of the Bayesian classifier. During testing of the PNN and its adaptive version using LVQ clustering technique [41], the LVQ classifies an input vector

by assigning it to the same class as the output unit which has its weight vector (codebook) closest to the input.

#### 4.1.1 LVQ 1

LVQ1 utilizes a straightforward procedure which suggests updating the weight towards the new input vector ( $x_i$ ) if the input and the weight vector belong to the same class or else updating the weight away from the input if the input and the weight vector belong to a different class (determined by finding the output pertaining to minimum distance i.e.,  $\|x_i - w_j\|$ ).

#### 4.1.2 LVQ2

This variant of LVQ involves updating the weights for the runner up distance subject to the condition that the ratios of runner up and closest distance i.e.,  $d_r/d_c > (1 - \epsilon)$  and  $d_c/d_r < (1 + \epsilon)$  ( $\epsilon$  is the window describing the error in the variance) with the additional constraint 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 of the input then updating the  $d_r$  and  $d_c$  are swapped. When the target is the nearest codebook then the updating of weight for that particular exemplar is not carried out.

#### 4.1.3 LVQ 3

A further improvement in procedures for updating the weight vector enables a mechanism for 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 ranging between 0.1 to 0.6 with smaller values corresponding to a narrower window.

### 4.2 Versions Of K- Means Clustering Algorithms

K-means algorithm finds the 'c' mean vectors ( $\mu_1, \mu_2, \mu_3, \dots, \mu_c$ ) where 'c' is the number of cluster centers. Hence, this elementary but popularly used clustering algorithm in common practice is called K-means algorithm. Several types of K-means algorithms have been developed over the years such as Lloyd (or) Forgy's K-Means, McQueen K-means, Standard K-means, Continuous K-Means, 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/ LLOYD'S K-MEANS

The algorithm describing this method is depicted in Fig. 2 below.

<u>Forgy's K- Means Algorithm:</u>
Step 1: Choose a set of random 'P' feature vectors and No. of Classes 'M'
Step 2: Select the first 'M' of 'P' feature vectors as initial prototypes
Step3: Assign each of the feature vectors to the nearest prototype to generate M classes:
A. Apply Index $c p =m$ to designate $x^{(p)}$ belongs to Class $m$ ;
B. Count Cluster Size with $s[m]$ ; Increment ( $s[m] = s[m] + 1$ ) go to Step 3
Step 4: Obtain the new centers in each class by averaging the feature vectors:
A. Initialize: $a[n][m] = 0$ where 'n' is the component of class 'm' and $x[n][p]$ is the n <sup>th</sup> component of $x^{(p)}$
B. For each Class 'm' initialize the averages
C. For the given Class 'm' find all the vectors in Cluster 'm'.
D. For each component 'n' obtain the sum the component values
E. Obtain the Mean Cluster 'm'
Step 5: IF [iterative process is not concluded] OR [no change in class] THEN go to Step 6 ELSE Step 3.
Step 6: Finish

Fig. 2: Generic Forgy's K-Means Algorithm

#### 4.2.2 Standard K-Means

Standard K-means algorithm [45] differs from the Forgy's algorithm in its better utility of the data at each stage. The generic process for both algorithms are the same in its context of choice of reference points as all points are assigned to clusters and the procedure uses the cluster centroids as reference in subsequent partitioning. However in this method the centroids are adjusted both during and after each partitioning process. For instance, for a data 'y' in cluster 'i' if the centroid  $z_i$  is the nearest reference point then no modifications are made and the algorithm proceeds to the next data. However, if the centroid  $z_j$  of the cluster 'j' is the reference point closest to the data 'y', then y is reallocated to cluster j, the centroids of the 'losing' cluster 'i' (minus point x) and the 'gaining' cluster (plus point x) are recomputed and the reference points  $z_i$  and  $z_j$  are moved to their new centroids.

### 4.3 Expectation Maximization With Maximum Likelihood Estimation

Expectation Maximization (EM) algorithm is an iterative procedure that utilizes the expectation process (E-step) accompanied by the maximization process (M-step) to obtain the Maximum Likelihood (ML) estimate in large feature sets that are usually accompanied by noise and missing data. The EM algorithm is guaranteed to converge to an ML estimate [46-47], and the convergence rate of the EM algorithm is usually reasonably quick. Each E-step computes the expected value of a set of ‘unobserved’ data by means of the recent parameter estimate and the observed data. Each M-step uses the value obtained from the E-step supposing it to be the measured data for obtaining a likelihood function and in turn in determining an ML estimate of the parameter. In both the steps an iterative procedure is pursued until the change in the log posterior likelihood function is minimized.

From the perspective of training the OPNN version, it is pertinent to note that in each E-step and M-step, the mean is constantly tuned until the log posterior likelihood function is minimized. Since the focus is on obtaining the estimates for the mean of Gaussian distribution of the training datasets, the EM algorithm utilizes a weight parameter, which is also updated during the computation of each step. The E step uses the PDF estimated in the second layer of the PNN along with the intra-class mixing coefficient ( $\beta$ ) to estimate the weight parameter. The M-step uses the data estimated in the E-step and the weight parameter  $w_{E,i}$  to form a likelihood function and serves in acquiring the ML estimate. Hence, using the weight calculated from the E step, the new values for the cluster center ( $C_{E,i}$ ) and the intra-class mixing coefficients ( $\beta_{E,i}$ ) are computed. Fig. 3 shows the EM algorithm implemented for obtaining the ML estimates for the PNN versions.

### 5. EXPERIMENTAL TEST SETUP AND PD PATTERN RECOGNITION STUDIES

Detailed laboratory studies and analysis have been carried out using a 10kVA, 100kV, 50Hz test transformer with a Digital PD Measurement and Acquisition System (W.S Test Systems make, DTM-D ) measurable in the range 2-5000pC. The PD system houses a Tektronix oscilloscope (TDS 2002B) with a tunable filter-insert (Model: DFT-1) with a variable center frequency (600 kHz- 2400 kHz) at a bandwidth of 9 kHz suitable for discharge measurement and data acquisition. Facility is provided to display the measured PD pulses either

in pico-coulomb (pC) or in milli-volt (mV) in accordance with stipulations laid down in IEC 60270 [48]. The straight detection and measurement test configuration as recommended in IEC 60270 is utilized in carrying out the studies since the experimentation has been carried out in a controlled laboratory environment thus precluding the need for alternative strategies for noise suppression albeit, the PD measurement and acquisition system comprises window gating facility to mask and mitigate unwanted background noise during measurement. Further, a 1nF coupling capacitor is also included in the test setup to improve the transfer characteristics of the test circuit. Calibration of the test setup is carried out using a reference calibrator (Model: PDG<sup>®</sup>) and in line with the requirements of IEC 60270.

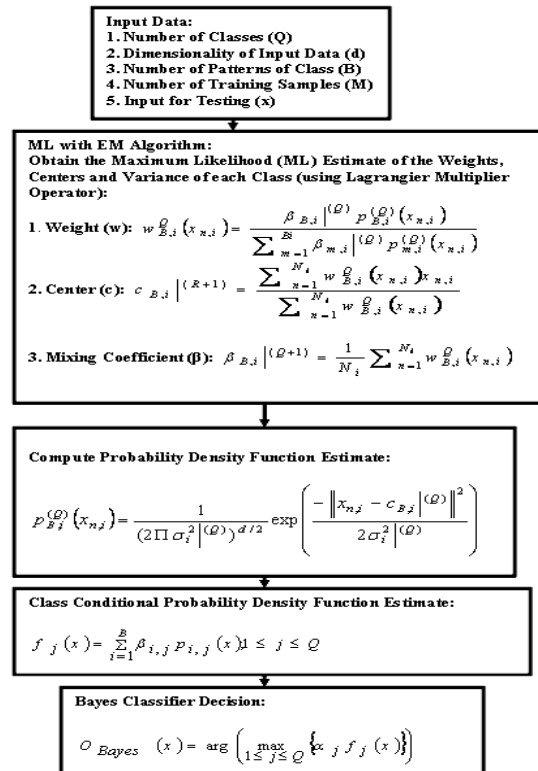


Fig. 3: Flow Chart of ML-EM Algorithm for Training OPNN

PD Gold<sup>®</sup> (HV Solution Inc.) data acquisition package is interfaced with the PD measurement system for acquiring the PD patterns. The package provides an opportunity to capture high resolution discharge signals at a sample rate of 1 sample per 2.5 nanoseconds (ns). In addition, the package also distinguishes PD for power frequency signals (50Hz) which enables the test engineer to study and

analyze the shape of the discharge pulses detected in addition to acquiring the phase resolved PD (PRPD) patterns in real-time. An additional feature of this software is that it provides an option of utilizing threshold level for recording the number of PD pulses (count) per cycle. The set-up process of this package during testing necessitates filling a customer-friendly 'Data Input Form' with easy to use dropdown menus. Provision is also made in the software for enhanced display of discharge pulses (both sinusoidal and elliptical time-base) with an option to view in auto or manual mode. In manual mode, the user is provided with the option for recording the discharge pulses for a substantial period. However, for the purpose of this research, the data acquired is limited to duration of 5-10 minutes after ensuring its appropriateness in line with stipulations for preconditioning and material properties of the test specimen [49]. Fig. 4 and Fig. 5 depict the PD measurement and acquisition system used in this study and a typical output of PD pulses acquired during the testing, measurement and acquisition process.

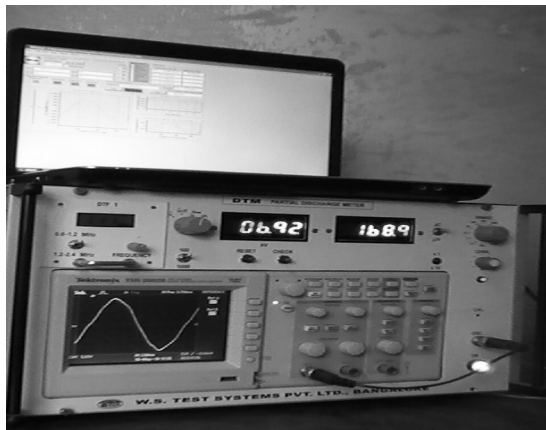


Fig. 4: PD Measurement and Acquisition System

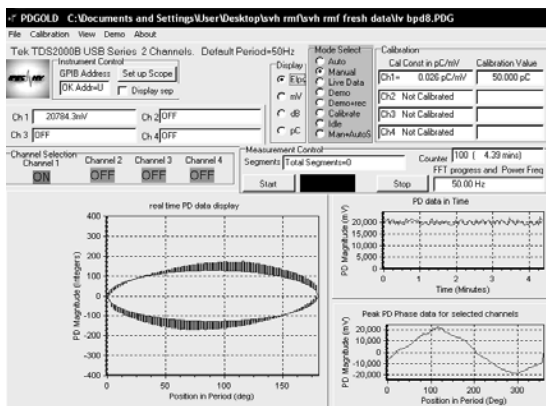


Fig. 5: Typical Patterns of Void with Corona PD

### 5.1 Pd Experimentation And Laboratory Models For Multiple Source Pattern Recognition

In order to ascertain the classification capability of the proposed PNN versions with and without clustering algorithms three exhaustive case studies have been taken up for detailed analysis. The analysis pertains to studies on the performance of the PNN versions with clustering algorithms for small, moderate and large datasets. The studies comprise three benchmark models pertaining to single source discharge patterns i.e., electrode bounded cavity (EC or V), air corona (C) and oil corona (OC) while three other unique models involve major forms of complex multiple source PD (overlapped patterns) namely the electrode bounded cavity with air corona (ECC), multiple electrode bounded cavity (MEC) and electrode bounded cavity with surface discharge (VC).

### 5.2 Benchmark Laboratory Models Replicating Single And Multiple Source Pd Patterns

Benchmark models are fabricated to replicate PD patterns that are archetypes of the source of discharge as recommended in [29]. Corona discharge in air is replicated with a point electrode initiated from a stainless steel rod with an approximate angle of 85° and a plane electrode 10mm thick, 60mm diameter. A similar arrangement immersed in transformer oil replicates the oil-corona form of discharges. Fig. 6 and Fig. 7 depict the models.

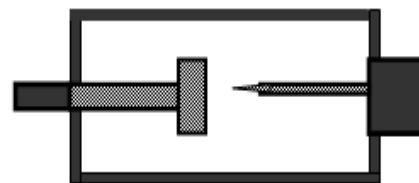


Fig. 6: Laboratory Benchmark Model simulating Air Corona

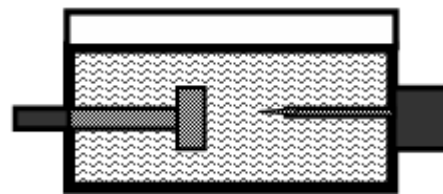


Fig. 7: Laboratory Benchmark Model replicating Oil Corona

12mm thick, 20 mm diameter perspex with electrode-bounded cavity of 2 mm depth simulates



internal PD. Electrode bounded cavity with air-corona is produced by inserting a needle configuration (2 mm) from the high voltage (HV) electrode in addition to a 2 mm electrode bounded cavity on perspex in the HV electrode. Multiple electrode bounded cavities replicate with nine numbers of 2mm and 4mm holes. Fig. 8 and Fig. 9 portray the arrangement utilized for analysis. The typical test setup representing the test arrangement is shown in Fig. 10.

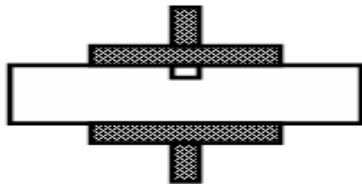


Fig. 8: Laboratory Benchmark Model representing Electrode Bounded Cavity

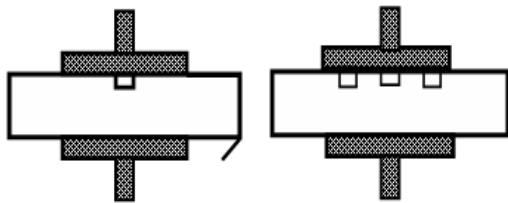


Fig. 9: Laboratory Models simulating Multiple-Source PD

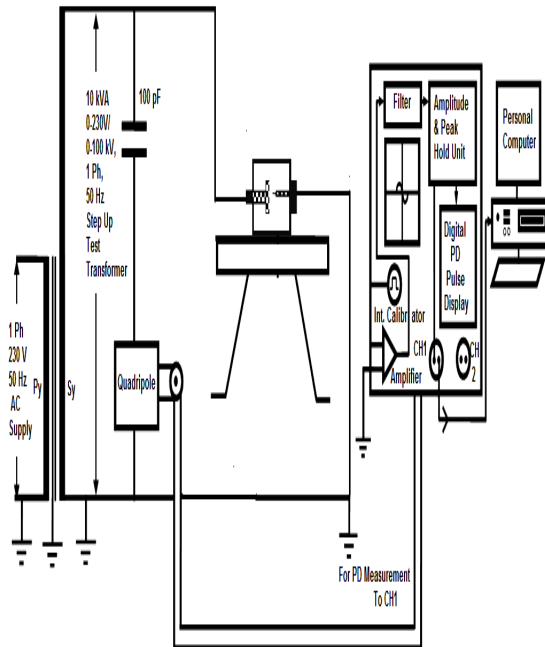


Fig. 10: Generic Layout of Test Setup for PD Pattern Recognition Studies (Air Corona Discharge)

The first case study comprises two sets of training data i.e. 20 and 25 sets. A total of 52 PD fingerprints samples were collected from 5 samples of laboratory benchmark models described in the previous discussions 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 8 patterns belong to electrode bounded cavity with air-corona (multi-source PD). The database obtained is indicated in Table 1.

Table 1: PD Signatures related to Small Dataset for Benchmark Laboratory Models

PD Class	Type of PD	Label for Classification	PD Patterns
1	Electrode Bounded Cavity	V	8
2	Surface Discharge	S	10
3	Oil-Corona	OC	10
4	Air-Corona	C	6
5	Electrode Bounded Cavity with Air-Corona	VC	8
6	Electrode Bounded Cavity with Surface Discharge	VS	10

The second case study involves a total of one hundred and twenty (120) PD signature patterns of each type of the defect for varying applied voltages. It is worth noting in this context that the patterns acquired exhibit the statistical variations in the pulse patterns for each cycle of the sinusoidal voltage thus displaying the inherent non-Markovian nature of PD making the task of classification more difficult. The task becomes even more demanding due to variations in applied voltages. Rigorous analysis is carried out to ascertain the classification capability of the modular PNN versions for each category of PD. Table 3 shows the patterns acquired for large dataset from various sources of PD. The third case study involves more samples of datasets during training and testing phase of OPNN for the PD signatures of models indicated in Table 2.

## 6. PREPROCESSING AND FEATURE EXTRACTION

Pre-processing and feature extraction ensures compactness of the input without compromise on

the distinctiveness of features in the raw data. The major objective of using various methods of feature extraction is to ascertain the performance of the NN in classifying patterns so that perceptible conclusions may be arrived at on the role played by the various key parameters of the network such as trainable part (smoothing parameter), curse of dimensionality and mean value of the probability density estimate etc. For extensive effective verification of training of the parsimonious versions of PNN, the features characterizing the data are created based on the phase window concept. Analysis is carried out utilizing measures related to simple statistical operators namely maximum/minimum values of a specific parameter, dispersion and central tendency. In addition, the capability of this scheme in addressing issues concerning curse of dimensionality during classification is analyzed and compared with the scheme involving simple statistical operators. In this context it is pertinent to note that prior to the real task of classifying PD patterns, the proposed PNN versions with their respective clustering algorithms were verified for their performance in classifying the classical Fischer's Iris Plant database.

Table 2: Large Database PD Signatures for several benchmark PD sources for Varying Applied Voltages

PD Class	Class Label	PD Type	Applied Voltage (kV)	No. of Patterns
1	EC	<b>Single Source:</b> Electrode Bounded Cavity	7.28	120
			9.1	
			9.6	
2	C	<b>Single Source:</b> Air-Corona	13.65	120
			20.93	
			22.75	
3	OC	<b>Single Source:</b> Oil-Corona	20.93	120
			29.12	
			31.85	
4	ECC	<b>Multiple Source:</b> Electrode Bounded Cavity with Air-Corona	9.1	120
			9.6	
			13.6	
5	MEC	<b>Multiple Source:</b> Multiple Electrode Bounded Cavity	7	120
			10	
			13	

## 7. OBSERVATIONS, ANALYSIS AND INFERENCES

Based on the testing of PNN and its adaptive version with two sets of training data which include single and overlapped PD source patterns comprising 4 sets (3 Nos. single PD source and 1 No. void-corona overlapped) and 5 sets (3 single PD sources and 2 Nos. void-corona and void-surface discharge overlapped) extensive observations and analysis are summarized.

1. OPNN has a major limitation in that it is fundamentally an unsupervised learning scheme (no updating of weights during the training phase). However, it is evident that more the representative sets of exemplar vectors during training phase the better is the ability of OPNN in classifying patterns. It is also pertinent to note that since the variation of variance parameter ( $\sigma$ ) is on a global basis i.e., all categories have the same value of smoothing parameter a fixed value of the parameter is utilized for the purpose of this study. The classification capability is summarized in Table 3.

2. It is also apparent during the analysis that since aspects related to over-fitting is a vital issue during training large non-Markovian PD datasets, the OPNN suffers from the drawback of requirement of large memory during the training phase thus indicating the need for clustering algorithms.

3. It is also evident from Table 3 that comparatively, the adaptive version in most cases is able to classify more types of defects during training phase (this network architecture is also not a supervised learning scheme) due to the inherent strength of the topology and configuration of the network (due to the individual and unique values of the variance parameter related to class separation).

4. Nevertheless, it is pertinent to note that due to inherent nature of being essentially an unsupervised scheme, greater misclassifications are observed in the case of multiple source PD patterns specifically with void-surface overlapped patterns. This aspect is substantiated as the characteristics of discharge pulses for multiple-source PD are extremely similar and almost fully overlapped imposing the need for training the network. This feature is also evident from Table 5 which clearly elucidates the peculiarity of the optimal decision hyper surface boundary with various values of sigma for each type of input.



Table 3: Comparison of Classification Capability of OPNN and APNN for small signature PD datasets without clustering

Feature Vector	Classification Rate (%) of OPNN		Misclassifications in OPNN		Classification Rate (%) of APNN		Misclassifications in APNN	
	4 Types	5 Types	4 Types	5 Types	4 Types	5 Types	4 Types	5 Types
$\Phi-q_{\max}^{-n}$ (30°)	90	80.8	6 Nos. : C2,C7, V1C2, V2C7, V5C2, V6C7	14 Nos. : C2,C7,V1C2, V2C7, V3C8, V4C9, V2S2, V3S3,V4S4, V1,V6C7, V6S6, V7S7,V8S8	93.3	91.78	4 Nos.: V1,V5, S6,S9	6 Nos.: V1, V6C7, V6S6, V7C8, V7S7, V8S8
$\Phi-q_{\min}^{-n}$ (30°)	88.3	80.8	7 Nos.: V5C2,V3C 8,V2C7, 1C2,V6C7, V7C8, V8C9	14 Nos.: V1C2,V2C7,V 3C8,V5C2,V1 S1,V2S2,V3S3 V4S4,V5S5, V1,V6C7,V6S 6,V7S7,V8S8	93.3	91.78	4 Nos.: V1,V6C 7,V7C8, V8C9	6 Nos.: V1,V6C7, V6S6, V7C8, V7S7, V8S8
$\Phi-q_{\max}^{-n}$ (10°)	90	89	6 Nos.: C2,C7,V1C 2,V5C2, V6C7, V8C9	8 Nos.: V5S5,V4S4, V1,V6C7, V7C8,V8C9, V7S7	93.3	94.5	4 Nos.: V2,V1, V3V5, S6,S7	4 Nos.: V1,V5, S6,S9
$\Phi-q_{\min}^{-n}$ (10°)	85	83.6	9 Nos.: C2,C7,V3C 7,V6C2,V8 C9,V3C8,V 2,V4,V6	12 Nos.: C2,C7,V3C7, V1C2,V6S6,V 7S7,V8S8,V6 C2,V2,V4	90	93.2	6 Nos.: S6,S7, S10, V6C2, V8C9	5 Nos.: V7S7,V6C 2,S10, S7,S6
$\Phi-q_{n_{\max}}$ (10°)	90	80.8	6 Nos.: V6C2, V8C9,C2, C10,V1C2, V3C8	14 Nos.: V2,V4,C2,C10 ,V1C2,V3C8, V5S5,V1,V3, V6C2,V8C9, V6S6,V7S7, V8S8	90	79.5	6 Nos.: V1,S3,S 4,S7,S9, V6C2	15 Nos.: V8,V7,V2, V4,V6,V1, V3,V5,C8, V1C2,V3C 8,V8C9,V6 S6,V7S7

**7.2 Performance Of Opnn And Apnn For Small Dataset Pd Signatures With Lvq Clustering Algorithms**

1. In almost all LVQ versions, it is observed that fewer misclassifications occur in the training phase of multiple-source PD pattern recognition. Only with the study of measures based on minimum and maximum values (with reference to the number of pulses) it is evinced that there are several misclassifications (about 7 numbers) of multiple source PD especially void-surface discharge. Exhaustive observations clearly indicate this aspect as shown in Table 4.

2. Further, it is worthwhile to note that this feature is made evident from the analysis indicated in Table 4 for various values of variance. This may also be

seen from the viewpoint of lower signal to noise ratio based data which in turn throws light on the possible anomalies during data acquisition in classification of PNN during training and testing of input database.

4. A significant aspect which is also apparent from Table 6 is the similarity in the range of values of variance for various classes based on each type of input with the exception of measures based on maximum and minimum values of number of pulses (10°). Incidentally the values of variance parameter in the case of APNN are found to only marginally different thus indicating that both the Bayesian Networks are able to create almost similar hyper surface boundaries. This aspect is made evident from the classification capability of PNN versions with LVQ algorithms.



Table 4: Comparison of Classification Capability of OPNN and APNN for small signature PD datasets with LVQ Versions of Clustering Algorithms

Feature Vector	Classification Rate (% of OPNN with LVQ 1)		Misclassifications in OPNN with LVQ1		Classification Rate (% of OPNN with LVQ 2)		Misclassifications in OPNN with LVQ2		Classification Rate (% of OPNN with LVQ 3)		Misclassifications in OPNN with LVQ3	
	4 Types	5 Types	4 Types	5 Types	4 Types	5 Types	4 Types	5 Types	4 Types	5 Types	4 Types	5 Types
$\Phi-q_{max}n(30^{\circ})$	95	91.8	3 Nos.: C2,C9, V1	6 Nos.: V1,V3, V7S7, V8S8, V2,V4	96.7	95.9	2 Nos.: V2,V1	3 Nos.: V1, V6S6, V8S8	96.7	91.8	2 Nos.: V2,V1	4 Nos.: V2,V1, V7S7, V8S8
$\Phi-q_{min}n(30^{\circ})$	93.3	91.8	4 Nos.: C7,C5, C2, V6C7, V7C8, V8C9	6 Nos.: V6C7, V7C8, V7S7,V8 S8,C2, V5C2	95	89.5	3 Nos.: V1,V6S6 V8S8	8 Nos.: C2,V5, S7,V6C7 V7C8, V6S6, V8S8	91.7	89.5	5 Nos.: C2,V5, V6C7, V7C8, V8C9	8 Nos.: C2,V5, V6C7, V7C8, V8C9
$\Phi-q_{max}n(10^{\circ})$	95	90.4	3 Nos.: V2, V8S8,V1	7 Nos.: V2,V1C 2,V5C2, V1,V3, V7S7, V8S8	98.3	93.2	1 No.: V1	5 Nos.: V1C2, V6C2, V6S6,V7 S7,V8S8	95	94.5	3 Nos.: V2, V5C2, V1	4 Nos.: V1C2, V6C2, V7S7, V8S8
$\Phi-q_{min}n(10^{\circ})$	96.7	94.5	2 Nos.: C2,C7	4 Nos.: C2,C7, V7S7, V8S8	96.7	93.2	2 Nos.: V5S5, V6C2	5 Nos.: V1C2, V6C2,V6 S6,V7S7, V8S8	98.3	94.5	1 No.: C2	4 Nos.: V1C2, V6C2, V7S7, V8S8
$\Phi-q_{max}n(10^{\circ})$	90	86.3	4 Nos.: V3C8,V 1,V6C2, V8C9	10 Nos.: V2,V4, V1C2, V1,V3, V6C2, V8C9, V6S6, V7S7	96.7	97.3	2 Nos.: V1C2, V6C2	9 Nos.: V4,V2S2 ,V5S5,V 3,V6C2, V8C9,V6 S6,V7S7, V8S8	90	84.5	4 Nos.: V1,V2, V6C2, V8C9	11 Nos.: V2,V4, V3C8, V1, V3, V6C2, V8C9, V6S6,V7 S7,V8S8
$\Phi-q_{min}n(10^{\circ})$	96.7	87.7	2 Nos.: C2,C7	9 Nos.: C2,C7, V2S2,V3 S3,V4S4 V5S5, V6S6, V7S7, V8S8	90	86.3	4 Nos.: C2,C7, V2C4, V7C4	10 Nos.: C2,V2C4 ,V2S2,V 3S3,V4S 4,V5S5V 7C4,V6S 6,V7S7, V8S8	98.3	89.1	1 No.: C2	8 Nos.: C2,V2S2 V3S3, V4S4, V5S5, V6S6, V7S7, V8S8

7.3 Comparison Of The Role Of Variance Paramter In Unsupervised And Supervised Pnn Verisons

1. It is pertinent to note from Table 5 that in the case of all the versions of LVQ-PNN versions, the range of the smoothing parameter which describes the feature for void defect is from 0.01 to 0.05. Similarly the value of  $\sigma_4$  i.e. void-corona overlapped pattern is also reasonably similar but for one specific case with LVQ3 only. This substantiates the fact which has already been stated by researchers in identifying and classifying the overlapped void-corona patterns. In addition, from the standpoint of decision hyper-boundary, considerable clarity in terms of separation of class boundaries is noticed.

2. However, in the case of void-surface fully overlapped patterns, the value of variance parameter is considerably atypical in various versions of LVQ. This feature is noticeable in the case of input feature vector using measures based on minimum and maximum values of number of pulses. This also reiterates the fact observed by several researchers studying PD pattern recognition.

3. This technique may prove to be effective when different sets of training data with larger training size are used and wider class identification is expected. This hence possibly suggests the need for more training for obtaining enough number of representative codebook vectors pertaining to a class for better class discrimination.



Table 5: Comparison on the Role of Variance Parameter in Classifying Multiple PD Source

Input Feature Vector		APNN (without clustering)	OPNN with LVQ1	OPNN with LVQ2	OPNN with LVQ3
$\Phi$ - $q_{\max}$ - $n$ ( $10^\circ$ )	4 Input Type	$\sigma_1=0.0969$ ; $\sigma_2=0.19539$ ; $\sigma_3=0.057904$ ; $\sigma_4=0.078573$ ;	$\sigma_1=0.018042$ ; $\sigma_2=0.025317$ ; $\sigma_3=0.070436$ ; $\sigma_4=0.020333$ ;	$\sigma_1=0.01884$ ; $\sigma_2=0.021165$ ; $\sigma_3=0.072821$ ; $\sigma_4=0.024945$ ;	$\sigma_1=0.033361$ ; $\sigma_2=0.040017$ ; $\sigma_3=0.061909$ ; $\sigma_4=0.016268$ ;
	5 Input Type	$\sigma_1=0.17094$ ; $\sigma_2=0.241086$ ; $\sigma_3=0.0669388$ ; $\sigma_4=0.0780457$ ; $\sigma_5=0.1734184$ ;	$\sigma_1=0.02412$ ; $\sigma_2=0.025053$ ; $\sigma_3=0.079018$ ; $\sigma_4=0.031718$ ; $\sigma_5=0.015538$ ;	$\sigma_1=0.0366$ ; $\sigma_2=0.0074722$ ; $\sigma_3=0.038068$ ; $\sigma_4=0.027677$ ; $\sigma_5=0.0039404$ ;	$\sigma_1=0.031258$ ; $\sigma_2=0.0077678$ ; $\sigma_3=0.040226$ ; $\sigma_4=0.029185$ ; $\sigma_5=0.003048$ ;
$\Phi$ - $q_{\min}$ - $n$ ( $10^\circ$ )	4 Input Type	$\sigma_1=0.1378$ ; $\sigma_2=0.206325$ ; $\sigma_3=0.070626$ ; $\sigma_4=0.075091$ ;	$\sigma_1=0.043939$ ; $\sigma_2=0.015394$ ; $\sigma_3=0.03495$ ; $\sigma_4=0.010627$ ;	$\sigma_1=0.038126$ ; $\sigma_2=0.020694$ ; $\sigma_3=0.043458$ ; $\sigma_4=0.019339$ ;	$\sigma_1=0.053615$ ; $\sigma_2=0.008649$ ; $\sigma_3=0.027969$ ; $\sigma_4=0.007127$ ;
	5 Input Type	$\sigma_1=0.17225$ ; $\sigma_2=0.2579$ ; $\sigma_3=0.008828$ ; $\sigma_4=0.09387$ ; $\sigma_5=0.14067$ ;	$\sigma_1=0.03234$ ; $\sigma_2=0.006136$ ; $\sigma_3=0.037399$ ; $\sigma_4=0.026545$ ; $\sigma_5=0.002986$ ;	$\sigma_1=0.0366$ ; $\sigma_2=0.007472$ ; $\sigma_3=0.03068$ ; $\sigma_4=0.027677$ ; $\sigma_5=0.0039404$ ;	$\sigma_1=0.031575$ ; $\sigma_2=0.006961$ ; $\sigma_3=0.038936$ ; $\sigma_4=0.027894$ ; $\sigma_5=0.002916$ ;
$\Phi$ - $q$ - $n_{\max}$ ( $10^\circ$ )	4 Input Type	$\sigma_1=0.136752$ ; $\sigma_2=0.192871$ ; $\sigma_3=0.0535503$ ; $\sigma_4=0.0624375$ ;	$\sigma_1=0.023925$ ; $\sigma_2=0.053411$ ; $\sigma_3=0.082635$ ; $\sigma_4=0.012081$ ;	$\sigma_1=0.016642$ ; $\sigma_2=0.05658$ ; $\sigma_3=0.09683$ ; $\sigma_4=0.013938$ ;	$\sigma_1=0.014868$ ; $\sigma_2=0.04014$ ; $\sigma_3=0.075515$ ; $\sigma_4=0.0095092$ ;

7.4 Performance Of Opnn And Apnn For Moderate Datasets With Labelled & Unlabelled Clustering Algorithms

1. It is evident during studies that the smoothing parameter becomes small which is attributed to the nature of the denseness of the density estimates of the PD pulses in the case of both partially and fully overlapped patterns which indicate the tendency of the peaked nature of the values. This explains the difficulties pertaining to classification of fully overlapped multiple PD sources.

2. It is appropriate to note that the versions of LVQ algorithms are able to create a reasonably good parsimonious set of centers relevant to the PD classes even with about 20% of prototype vectors. It is to be emphasised that these codebook vectors become the weight (centers) vectors which are now the representatives of the samples.

3. It is also obvious from Table 6 that the LVQ 2 version of clustering algorithm out performs the rest of the types in classifying multiple PD patterns. A few researchers in other fields have also concurred on similar observations for large dataset training and testing utilizing LVQ 2 version.

4. It is also evident from Table 6 that there is a noticeable level of classification capability of patterns pertaining to the feature extraction scheme that utilizes the inequality relation based on measures related to the types of mean values (both 30° and 10° phase window input features).

5. It is observed from Table 6 that the classification capability of the unlabelled clustering algorithms though for the same quantum of cluster representatives (only 20% of the training datasets) is quite inferior as compared to the labelled algorithms and also as concurred by researchers in the field. The ability of such algorithms to provide class separable boundaries, offers an interesting prospect in carrying out input data validation in addition to providing plausible solutions for identifying unknown categories.

7.5 Performance Of Pnn Using Em-Ml Algorithm For Large Dataset Classification

To ascertain the capability of the PNN in classifying large dataset multiple source PD, the dataset is trained fundamentally with two different sets of PD signatures that are representative of the discharge source namely: 1. 360 sets (90 sets from each class) pertaining to patterns from 4 classes of PD sources (electrode bounded cavity-‘EC’, air-corona-‘AC’, oil-corona-‘OC’ and electrode bounded cavity with overlapped air corona-‘ECC’) and 2. 450 sets (90 sets of each class) involving 5 classes of PD sources (electrode bounded cavity-‘EC’, air-corona- ‘AC’, oil corona-‘OC’, electrode bounded cavity with air-corona- ‘ECC’ and multiple source electrode bounded cavity- ‘MEC’).



Table 6: Classification Capability of OPNN and APNN with LVQ and K-Means Clustering Algorithms

Input Feature Vector	Classification Capability																									
	LVQ1 Clustering				LVQ2 Clustering				LVQ3 Clustering				Standard K Means				Forgy's K Means									
	OPNN		APNN		OPNN		APNN		OPNN		APNN		OPNN		APNN		OPNN		APNN		OPNN		APNN			
	Iter.	(%)	Iter.	(%)	Iter.	ε	(%)	Iter.	ε	(%)	Iter.	ε	η	(%)	Iter.	ε	η	(%)	Iter.	(%)	Iter.	(%)	Iter.	(%)		
$\Phi$ - $q_{0.001}$ -n (30)	5000	89	5000	91	500	.5	95	1000	.5	94	1000	.6	.3	93	1000	.7	.3	93	1000	81	1000	83	5000	80	5000	81
$\Phi$ - $q_{0.01}$ -n (10)	5000	87	5000	88	1000	.6	90	1000	.7	93	1000	.8	.2	89	1000	.7	.2	89	1000	84	1000	84	5000	81	5000	82
$\Phi$ - $q_{0.01}$ -n (30)	1000	91	1000	93	1000	.6	94	1000	.8	93	5000	.7	.3	92	1000	.6	.3	93	5000	80	1000	81	5000	80	5000	81
$\Phi$ - $q_{0.01}$ -n (10)	1000	88	1000	89	1000	.7	91	1000	.7	93	1000	.7	.3	91	1000	.8	.3	92	5000	81	1000	82	5000	81	5000	81
Measure of Types of Mean (30)	1000	91	1000	93	1000	.8	95	1000	.8	96	1000	.7	.3	92	3000	.7	.3	93	5000	84	1000	85	5000	82	5000	82
Measure of Types of Mean (10)	1000	92	1000	93	1000	.8	95	1000	.8	96	1000	.7	.3	92	1000	.8	.3	94	5000	85	1000	86	5000	83	5000	83
Debauches Coefficients	1000	94	1000	94.5	1000	.5	96.5	1000	.6	97.5	1000	.6	.3	94.5	1000	.6	.3	95.5	2000	89	1000	91	5000	85	5000	87

Further it is also important to note that during the course of the study, a fixed value of the smoothing parameter ( $\sigma = 0.0001$ ) as a common variance value has been taken up for PNN training since it is observed that for almost all the feature extraction datasets the class decision boundaries are reasonably well deciphered.

During the training phase, randomly selected preprocessed vectors for the initialization of representative centers are taken. For the purpose of study and comparison Set 1 centers are picked up such that the vectors are alternatively odd and even numbered samples of PD signatures pertaining to each applied voltage. However, in order to check any marked change in the classification capability of the proposed PNN versions that are trained based on EM-ML algorithm in discerning the Set 2 centers pertaining to each applied voltage are sampled such that the last two signature vectors are taken up for study. Table 8 summarizes the performance of the proposed PNN versions.

The noteworthy aspects observed in Table 7 are summarized:

1. Misclassification rate invariably is about double for each type of randomly selected centers. It is also pertinent to note that the classification rate of the

PNN is about 1-2% better (about 10 to 15 lesser number of misclassified patterns) thus indicating the inherent capability of this version of PNN in creating independent class density estimates.

2. It is also evident from Table 7 that a more parsimonious set of centers is obtained utilizing the EM-ML algorithm with superior classification capability. However, problems related to the effect of outliers in discriminating inter-class clusters is observed to pose some difficulties during classification of patterns pertaining to relatively fully overlapped patterns (Class labels 'EC', 'AC' and 'ECC').

3. It is also obvious from Table 7 that the number of PDFs created is based on this algorithm with OPNN has divergent changes which may be attributed to the lack of the initialization in the input seed vectors that are representative to each class.

4. A variation of this scheme might be to utilize the Heteroscedastic PNN (HRPNN) which employs a different variance pertaining to classes. Studies carried out by the authors of this research [36] have clearly indicated the merits of utilizing the EM-ML algorithm for training large dataset multiple source PD patterns.

Table 7: Classification Capability of OPNN with EM- ML Algorithm for Large Datasets with randomly selected centers

Input Feature	No. of Tuples	Total No. of Testing Datasets	Randomly Chosen Initial Centers- 144 Nos. (36 Sets for each class)			Classification Capability of PNN with EM-ML Algorithm (%)		
			No. of PDF labeled sequentially as EC, AC, OC and ECC (Set 1 only)		No. of Iterations			
			OPNN		Set 1	Set 2	Set 1	Set 2
$\phi$ - $q_{\max}$ - $n$ (30°)	36	621	24,22, 20,21		28	16	81.33	79.8
$\phi$ - $q_{\min}$ - $n$ (30°)	36	621	22,25, 24,18		30	14	79.1	77.3
$\phi$ - $q_{\max}$ - $n$ (10°)	108	624	12,14, 15, 14		36	15	83.37	80.2
$\phi$ - $q_{\min}$ - $n$ (10°)	108	624	16,11, 16,12		43	27	81.7	78.6
Traditional Statistical Operators (30°)	48	621	14,16, 14,10		11	8	89.6	87.1
Traditional Statistical Operators (10°)	144	621	12,14, 18,11		9	6	91.1	88.2

## 8. CONCLUSIONS

It is evident from this research that the approach of utilizing clustering algorithms for PD pattern classification using various versions of PNN has the following distinct advantages:

1. This methodology offers fresh insight into solutions pertaining to issues related to on-line testing and the associated complexities as compared to previously attempted approaches including those carried out by the authors of this research earlier.
2. Though it is obvious that only the rudimentary clustering algorithms are utilized in this research, these algorithms (in particular the labelled LVQ versions) are able to handle the complexities pertaining to training large ill-conditioned dataset and overlapped multiple PD sources considerably well. This scheme may find applicability in training large datasets obtained during controlled sample testing to simulate a wide variety of flaws in insulation which are know a-priori.
3. The unlabelled clustering algorithm offers an interesting alternative to carry out cluster validation

and hence provides a plausible methodology for identification of unknown categories of flaws/ defects during PD evaluation. However, it is vital to note that the application being of critical importance (assessing insulation system and the level of risk associated with the operation of the power apparatus) it may be appropriate that the patterns of discharges are rather misclassified into a known category (as a strong learning scheme) and then builds on its capability to create classes based on further learning and validation which then clearly demarcates into an unknown category. This approach would be more suitable considering the criticality of the application rather than utilizing the weak learning strategy straightaway. The research studies by the authors of this work are progressing along the lines of proposed approach.

4. Notwithstanding the fact that training PNN versions utilizing the EM-ML algorithm serve as a reasonably good methodology for training the networks, it is evinced from the detailed analysis that the classification capability of the NN versions can be enhanced further by utilizing the clustering algorithms to obtain better initial cluster representatives. This is evident from the analysis

carried out even with the rudimentary versions of autonomous clustering algorithms. Research is under progress with regard to ascertaining the performance of much superior versions of clustering algorithms such as Global K-Means and Continuous K-means.

5. It is also clear from the study that though only the rudimentary clustering algorithms are utilized for obtaining the representative initial seed vectors in this research, these algorithms (in particular the deterministic version- LVQ 2) are able to handle the complexities pertaining to large ill-conditioned dataset training, difficulties pertaining to overlapped multiple PD sources and issues related to variations in patterns due to varying applied voltages considerably well. This scheme may find applicability in training large datasets obtained during controlled sample testing to simulate a wide variety of flaws in insulation which are known a-priori.

#### ACKNOWLEDGEMENTS

This research was supported by the Research and Modernization Fund (RMF) Grant, Project No.: 6, constituted by SASTRA University. The first author is extremely grateful to Prof. Sethuraman, Vice- Chancellor, SASTRA University Dr. S. Vaidhyasubramaniam, Dean- Planning & 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.

#### REFERENCES:

- [1] E. Lemke, Sonja Berlijn, Edward Gulski, Michael Muhr et al, "Guide for Electrical Partial Discharge Measurements in Compliance to IEC 60270", Technical Brochure 366, Electra, WG D1.33, CIGRE, No. 241, 2008, pp. 61-67.
- [2] 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, February 2000, pp. 40-47.
- [3] June-Ho Lee, Tatsuki Okamoto and Chin Woo Ti, "Classification of PD Patterns from Multiple Defects", Proceeding of 6<sup>th</sup> International Conference on Properties and Applications of Dielectric Materials, Xi'an Jiatong University, China, June 21-26, Vol. 1, 2000, pp. 463 – 465.
- [4] E. Gulski and A. Krivda, "Neural Network as a Tool for Recognition of Partial Discharges", IEEE Transactions on Electrical Insulation, Vol. 28, Issue 6, 1993, pp. 984–1001.
- [5] A. A. Mazroua, R. Bartnikas and M. M. A. Salama, "Discrimination between PD pulse shapes using different neural networks paradigms", IEEE Transactions on Dielectrics and Electrical Insulation, Vol.7, Issue 6, 1994, pp. 1119–1130.
- [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, Issue 2, 1994, pp. 265–275.
- [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, Issue 2, 2002, pp. 446-456.
- [8] M. Hoof, B. Freisleben and R. Patsch, "PD source identification with novel discharge parameters using counter propagation neural networks", IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 4, No. 1, 1997, pp. 17–32.
- [9] 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, 2005, pp. 2258- 2263.
- [10] A. Contin, A. Cavallini, G.C. Montanari and G. Pasini, "Digital Detection and Fuzzy Classification of Partial Discharge Signals", IEEE Transaction o Dielectrics and Electrical Insulation, Vol. 9, 2002, pp. 335-348.
- [11] L. Satish and W. S. Zaengl, "Can Fractal Features be used for recognizing 3-D Partial Discharge Patterns", IEEE Transaction on Dielectrics and Electrical Insulation, Vol. 2, No. 3, April 1995, pp. 352–359.
- [12] 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, August 1998, pp. 550-557.
- [13] 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, Issue 4, August 2004, pp. 715-723.





- [14] L. Satish and B. I. Gururaj, "Use of Hidden Markov Models for Partial Discharges pattern classification", IEEE Transactions on Electrical Insulation, Vol. 28, No. 2, 1993, pp. 178-182.
- [15] 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", Proceedings of 17<sup>th</sup> International Symposium on High Voltage Engineering (ISH 2011), 22<sup>nd</sup> - 26<sup>th</sup> August 2011, Hannover, Germany, F-039, pp. 401.
- [16] 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, February 2000, pp. 40-47.
- [17] 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, June 2002, pp. 446-457.
- [18] R. J. Van Brunt, "Stochastic Properties of Partial Discharge Phenomenon", IEEE Transactions on Electrical Insulation, Vol. 26, Issue 5, October 1991, pp. 902-948.
- [19] R. J. Van Brunt, E. W. Cernyar and P. Von Glahn, "Importance of Unraveling Memory Propagation Effects in Interpreting Data on Partial Discharge Statistics", IEEE Transaction on Electrical Insulation, Vol. 28, Issue 6, December 1993, pp. 905-916.
- [20] N.C. Sahoo, M.M.A. Salama and R. Bartinkas, "Trends in Partial Discharge Pattern Classification: A Survey", IEEE Transaction on Dielectrics and Electrical Insulation, Vol.12, No.2, April 2005, pp. 248-264.
- [21] M.G. Danikas, A.D. Karlis, "On the Use of Neural Networks in Recognizing Sources of Partial Discharges in Electrical Machine Insulation: A Short Review", International Review of Electrical Engineering (I.R.E.E.), Vol.1, No.2, June 2006, pp. 277-285.
- [22] CIGRE Working Group Report 21.03, Recognition of Discharges, Electra No. 11, 1969, pp. 61-98.
- [23] J. P. Steiner, "Partial Discharge - Part IV: Commercial PD Testing", IEEE Electrical Insulation Magazine, Vol. 7, No. 1, 1991, pp.20-33.
- [24] E. Gulski, "Digital analysis of partial discharges", IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 2, No. 5, 1995, pp. 822-837.
- [25] A. Krivda, "Automated Recognition of Partial Discharges", IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 2, No. 5, 1995, pp. 796-821.
- [26] R.E. James and B.T. Phung, "Development of Computer- based Measurements and their Application to PD Pattern Analysis", IEEE Transaction on Dielectrics and Electrical Insulation, Vol. 2, 1995, pp. 838-856.
- [27] R. Candela, G. Mirelli and R. Schifani, "PD Recognition by means of Statistical and Fractal Parameters and a Neural Network", IEEE Transaction on Dielectrics and Electrical Insulation, Vol. 7, 2000, pp. 87-94.
- [28] Mazroua, A. Amira and M. M. A. Salama, "PD pattern recognition with Neural Network using the Multilayer Perceptron technique", IEEE Transactions on Electrical Insulation, Vol. 28, No. 6, 1993, pp. 1082-1089.
- [29] N.B. Bish, P.A. Hawson, et al., "Combined Intelligent PD Analysis of High Voltage Dielectric Condition Evaluation", Proceedings of International Symposium on Soft Computing and Intelligent Systems for Industry (SOCO/ ISFI 2001), June 26<sup>th</sup>- 29<sup>th</sup>, UK, 2001.
- [30] B. Karthikeyan, S. Gopal and S. Venkatesh, "ART 2- An Unsupervised Neural Network for PD pattern recognition and classification", International Journal on Expert Systems with Applications, Elsevier, Vol. 31, 2006, pp. 345-350.
- [31] 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, IEE Centre, Slovakia, Vol. 57, No.3, 2006, pp. 138-145.
- [32] 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, World Scientific, Vol.2, No.4, December 2005, pp. 1-12.
- [33] B. Karthikeyan, S. Gopal and S. Venkatesh, "A Heuristic Probabilistic Neural Network System for Partial Discharge Pattern Classification", Journal of Indian Institute of Science, Bangalore, 85, 2005, pp. 279-294.
- [34] B. Karthikeyan, S. Gopal and S. Venkatesh, "Partial Discharge Pattern Classification Using



- Composite Versions of Probabilistic Neural Network Inference Engine”, Expert Systems with Applications, 34, 2008, pp. 1938-1947.
- [35] 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, No. 38, 2011, pp. 8978-8989.
- [36] 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, No. 38, 2011, pp. 11501-11514.
- [37] Donald F. Specht, “Probabilistic Neural Networks for Classification, Mapping or Associative Memory”, IEEE International Conference on Neural Networks, Vol.1, No.1, 1998, pp. 525-532.
- [38] Donald F. Specht, “Probabilistic Neural Networks and the Polynomial Adaline as Complementary Techniques for Classification”, IEEE Transactions on Neural Networks, Vol.1, No.1, March 1990, pp. 111-121.
- [39] C. H. Chen, “Fuzzy Logic & Neural Network Handbook”, 1<sup>st</sup> ed., McGraw-Hill, New York, 1996.
- [40] D. F. Specht and H. Romsdahl, “Experience with Adaptive Probabilistic Neural Networks works and Adaptive General Regression Neural Networks”, Proceedings of IEEE International Conference on Neural Networks, FL, Vol. 2, 1994, pp. 1203-1208.
- [41] N. K. Bose and P. Liang, “Neural Network Fundamentals with Graphs, Algorithms and Applications, McGraw- Hill Inc., Hightstown, NJ, 1996.
- [42] Vance Faber, “Clustering and the Continuous K- Means Algorithm”, Journal of Los Alamos Science, 22, 1994, pp. 38-144.
- [43] Rui Xu and Donald Wunsch II, “Survey of Clustering Algorithms”, IEEE Transactions on Neural Networks, Vol. 16, No. 3, May 2005, pp. 645-678.
- [44] Laurene Fausett, “Fundamentals of Neural Networks- Architectures, Algorithms and Applications”, Pearson Education, 2004.
- [45] Hyung Suck Cho, “Opto- Mechatronic Handbook: Techniques and Applications”, Chapter 9, Pattern Recognition, CRC Press, Florida, 2003.
- [46] R. L. Streit and T. E. Luginbuhl, “Maximum likelihood training of probabilistic neural networks”, IEEE Transactions on Neural Networks, 5, 1994, pp. 764–783.
- [47] Wu C. “On the convergence properties of the EM algorithm”, Annals Statistics, 11, 1983, pp. 95–103.
- [48] IEC 60270, High Voltage Test Techniques- Partial Discharge Measurements, 2000.
- [49] ASTM, D 61847, Practice for Conditioning Plastics for Testing, 2003.

**AUTHOR PROFILES:**



**S. Venkatesh** is an Assistant Professor in the Department of Electrical and Electronics Engineering, School of Electrical and Electronics Engineering, SASTRA University, India. He obtained a bachelor degree in electrical and electronics engineering from Annamalai University, India in 1994 and Master Degree in Technology-High Voltage Engineering from SASTRA, Deemed University in 2004. He is presently pursuing his Ph.D. in High Voltage Engineering. He was with Alstom Limited, Systems Group-India, T & D Projects Division, as assistant manager of engineering & QA. His specific area of interest includes Partial Discharge, Substation Engineering & EHV Transmission systems and Pattern Recognition and Artificial Intelligence.



**S. Gopal** is the Executive Director of WS Test Systems Limited, Bangalore, India. He obtained his MIE degree from Mysore University, India and his Master degree from Mysore University, India in 1978. He received a PhD from Indian Institute of Technology (IIT Madras), Chennai, India in 1984. He underwent advanced training in high voltage engineering in Germany for a period of one year. His specific area of research includes impulse testing of power transformers and partial discharge phenomena in power apparatus.



**S. Jayalalitha** is the Associate Dean in the Department of Electronics and Instrumentation Engineering, School of Electrical and Electronics Engineering, SASTRA University. She obtained her bachelor degree in Electronics and Communication Engineering in 1998 and Master Degree in Power Systems Engineering in 1993. She received a PhD from Indian Institute of Technology (IIT Madras), Chennai in Instrumentation in 2006. Her specific area of research includes High Voltage Measurement and Instrumentation techniques, Analog Circuits and Power System Harmonics.