



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJIRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 8, August 2021

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.569**

 9940 572 462

 6381 907 438

 [ijirset@gmail.com](mailto:ijirset@gmail.com)

 [www.ijirset.com](http://www.ijirset.com)

# A Literature Review of Methods (GNB, KNN and Xgboost) Applied for Identification of Partial Discharge Source

Ms. Kajol Chaudhari <sup>1</sup>, Dr. Akhilesh Kumar Mishra <sup>2</sup>, Dr. Priyanka Kothoke <sup>3</sup>,

Mr. Yogesh Chaudhari <sup>4</sup>

Ph.D. Research Scholar, Shri J. J. T. University, Jhunjhunu, Rajasthan, India<sup>1</sup>

Professor, Shri J. J. T. University, Jhunjhunu, Rajasthan, India<sup>2</sup>

Assistant Professor, MGM College of Engineering and Technology, Mumbai, Maharashtra, India<sup>3</sup>

Assistant Professor, School of Engineering and Technology, Navrachana University, Vadodara, Gujarat, India<sup>4</sup>

**ABSTRACT:** Partial discharge (PD) detection, measurement and classification constitute an important tool for quality assessment of insulation systems utilized in HV power apparatus and cables. The patterns obtained with PD detectors contain characteristic features of the source/class of the respective partial discharge process involved. The recognition of the source from the data represents the classification stage. Usually, this stage consists of a two-step procedure, i.e., extraction of feature vector from the data followed by classification/ recognition of the corresponding source. The various techniques available for achieving the foregoing task are examined and analyzed; while limited success has been achieved in the identification of simple PD sources, recognition and classification of complex PD patterns associated with practical insulating systems continue to pose appreciable difficulty.

**KEYWORDS:** Partial discharge, PD pulse, PD detector, pattern classification

## I. INTRODUCTION

Partial Discharge is a very common phenomenon occurring in electrical engineering insulation systems under high voltage (HV) stress[1]. It is caused due to localized dielectric breakdown bridging the gap between the conductor and insulation when the electrical field around is higher than the breakdown field of an insulating medium. PD is the fundamental reason for the ageing of electrical insulation and breakdown of electrical-mechanical assembly functioning at high voltage. This can even lead to a complete failure of the supply system as a whole. Hence, it cannot be ignored. Partial discharge can be prevented starting with suitable material selection and a thorough design[2]. It is recommended that all the equipments are tested in the manufacturing stage itself. It is mandatory to have periodical check-ups of the high voltage equipment to avoid long term failures for reliable operation. Thorough monitoring of Bushes is also very important in case of large power transformers as problems in bushing can easily pass undetected and occur unexpectedly.

Partial discharges can be of different gravities. Some discharges are relatively harmless like the corona into the air from outdoor cable setting ends. Whereas some discharges are extremely dangerous causing a great deal of damage to the health of the insulation system like the discharges within polymeric cables and accessories. During testing, a professional engineer determines the exact type of discharge and its origin so as to notify the potential risks involved. Various methods have been developed for detection of partial discharge in order to prevent complete system failure and have been fully acknowledged early in the last century. PD sensors can be used at places suggested during testing to get the relevant data for any high voltage equipment, which can then be verified for type of discharge or classification of discharge type[3].

Two vital conditions cause partial discharge between conductor and insulation as well as in insulation:

1. Presence of an initial electron to initiate an avalanche.
2. When the on line electrical field is higher than the original ionization field of the insulating medium

Thus, PDs are mainly concerned with dielectric materials utilized and connecting electrodes between which the voltage is applied. The insulation may comprise of solid, liquid, or gaseous materials, or any blend of them[4]. Various wellsprings of PD give diverse impact on insulation performance. The occurrence of sparks, flashes, arcs and electrical discharges are certain indications of existence of insulation problems. Subsequently, PD classification aims at the recognition of discharges of obscure origin and is significant to assess the level of the caused destruction[5].

## II. CLASSIFICATION OF PD TYPES

Initially, the PD classification was done by examining its patterns by naked eyes using the well-known ellipse, on an oscilloscope screen. However, it has been proved to be a crude method by latest research. In current scenario, the published research distinguishes the types of PD sources by using intelligent techniques like artificial neural networks, fuzzy logic, and acoustic emission[8]. With these sophisticated digital processing techniques, we gain exclusively new insight into the physical and chemical properties of PD phenomena and can characterize PD 'patterns' as:

1. Void indicated by a peak point,
2. Surface being all over the surface and
3. Corona being crown shaped.

These patterns in return enable the identification of the characteristics of the insulation 'defects' created by the PDs. Hence, this research enables to characterize the type of PD, though varying in its gravity, so as to remove particular insulation defects and avoid total supply failure[9][10].

The three distinct classifications of PD pulse data patterns gathered from the digital PD detectors during the experiments are -- phase-resolved data, time-resolved data and data having neither phase nor time information.

1. The phase-resolved data consists of three-dimensional discharge epoch,  $\phi$  charge transfer,  $q$  discharge rate,  $n$  patterns ( $\phi\sim q$ ,  $q\sim n$  and  $\phi\sim n$  patterns) at some specific test voltage.
2. The time-resolved data constitute the individual discharge pulse magnitudes over some interval of time, i.e.,  $q\sim t$  data pattern.
3. The third classification of information comprises of variations in discharge pulse magnitudes against the amplitude of the test voltage  $V$  (for both increasing and decreasing levels), i.e.,  $q\sim V$  data.

Out of these three classifications, phase resolved data method is used in this research work for all methods since it is drawing 2D patterns from raw data showing variations between either phase angle and charge, phase angle and number of pulses or number of pulses and charge and then comparison of these patterns amongst themselves will result in improved accuracy of the outputs.

At present, in any industry, just the source of PD is notified by using a buzzer. However, the buzzer indicates only the place of its occurrence and not the type of discharge, which is very essential for removal of insulation defects. The various techniques suggested in this research enable to know the type of the PD accurately so that the insulation defects can be removed from any high voltage equipment by specific techniques to avoid chances of supply failure.

## III. CURRENT METHODS USED FOR DETECTION OF SOURCE OF PD

Currently, the type of discharge is identified by Signal processing tools, Wavelet transform, Image processing tools or Distance classifiers and many more. The drawbacks of these methods are –

1. **Signal processing tool** extracts the PD patterns only with respect to frequency and not with respect to time. Hence, an effort is made to characterize the PD patterns/types directly in time domain for quick removal of insulation defects.
2. **Wavelet transform (WT)** has been applied to carry out time-frequency analysis in a number of problems. A continuous wavelet transform is generally used to obtain the time-varying information of the patterns by measuring similarity between the signal and a set of fast-decaying and oscillating functions called as wavelets. However, the time required by this method is much more to detect the type of discharge. Hence, an effort is made to apply two methods viz: Support vector machine method and Neural network method to quicken the detection of the PD.
3. **In Image processing tool**, the phase-resolved three-dimensional  $\phi$ - $q$ - $n$  pattern is regarded as an image and many modern image processing algorithms are used to extract the distinguishing features of the image.

However, this image as an input to the distance classifiers will not give accurate result about the PD pattern clusters and hence the insulation defects. This is because the distance classifier assigns the pattern to the class to which it is matching with the closest minimum distance but cannot compare with known patterns for sure shot results.

#### IV. IMPROVED METHODS USED IN CURRENT RESEARCH

This research improves all the above methods. Firstly, we need to find statistical parameters using Statistical method and give them as an input for various methods such as Gaussian Naive Bayes (GNB), K-Nearest Neighbour (KNN) and XGBOOST methods for accurate identification of discharge type, comparable with known patterns. The sequence of the methods applied in my research for detecting the type of PD is –

##### 1. Statistical methods

These methods are applied for phase-resolved data patterns obtained from raw data use standard univariate statistical analysis for the computation of several statistical parameters characterizing the univariate distributions,  $\phi$ -q,  $\phi$ -n and n-q. These univariate parameters  $\phi$ , q, n at some specific test voltage V constitute the feature vector. The five parameters obtained as a result of Statistical Analysis i.e. mean, standard deviation, variance, skewness and kurtosis for n-q and skewness and kurtosis for  $\phi$ -q and  $\phi$ -n are given as input for all four methods discussed below.

Paper [1] is the survey paper in which only various methods are listed for PD type identification, out of which Statistical method for detecting type of PD has been chosen in this research as first method. Paper [6, 7], explain methodology applied for the same. Paper [6] has listed several statistical indexes which can be used for PD type identification. Paper [7] explained about different statistical parameters and hitting patterns used for the identification of PD type of ac rotating machines. However, Hitting patterns follow conventional method of classification depending on “Recognition Rate.” The bar graphs called as Hitting Patterns obtained using statistical parameters cannot differentiate two different patterns of the same type of discharges having same recognition rate though from different physical conditions. Hence, to overcome this limitation this research uses same statistical parameters such as mean, standard deviation, variance, skewness and kurtosis for phase resolved partial discharge patterns between ( $\phi$ -q,  $\phi$ -n and n-q) in which the recognition rate cannot be same due to the plotting of two variant quantities [14 - 16]. Accuracy achieved by using this method is 85 %. To try to improve the accuracy of detecting the type of PD, will be implementing few more methods.

##### 2. Gaussian Naive Bayes (GNB) Method:

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. There is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the colour, roundness, and diameter features[11].

For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods.

Despite their naive design and apparently oversimplified assumptions, naive Bayes classifiers have worked quite well in many complex real-world situations. In 2004, an analysis of the Bayesian classification problem showed that there are sound theoretical reasons for the apparently implausible efficacy of naive Bayes classifiers. Still, a comprehensive comparison with other classification algorithms in 2006 showed that Bayes classification is outperformed by other approaches, such as boosted trees or random forests. An advantage of naive Bayes is that it only requires a small number of training data to estimate the parameters necessary for classification.

Naïve Bayes is a subset of Bayesian decision theory. It's called naive because the formulation makes some naïve assumptions. Python's text-processing abilities which split up a document into a vector are used. This can be used to classify text. Classifies may put into human-readable form. It is a popular classification method in addition to conditional independence, overfitting, and Bayesian methods. In the face of the simplicity of Naive Bayes, it can classify documents surprisingly well. Instinctively a potential justification for the conditional independence assumption is that if the document is about politics, this is a good evidence of the kinds of other words found in the document. Naive Bayes is a reasonable classifier in this sense and has minimal storage and fast training, it is applied to time-storage critical applications, such as automatically classifying web pages into types and spam filtering. Considering a set of objects, each of which belongs to a known class, and each of which has a known vector of variables, the aim is to create a rule which enables to allocate future objects to a class, given just the vectors of variables marking out the future objects. These problems are known as supervised classification problem, are worldwide, and most of the methods for constructing such rules have been developed. It is very easy to establish, and no need any complicated repetitive parameter estimation schemes. This means it should be applied to huge data sets. It is easy to interpret, so unskilled users in classifier technology can make out the reason for it is making the classification it makes. Finally, it often does surprisingly well: it should not be the best possible classifier in any particular application, but it can usually be relied on to be robust and to do well.

**3. K-Nearest Neighbour (KNN) Method:**

KNN is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

Example: Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category (as shown in Fig.1).



Fig.1. KNN model identification technique

Suppose there are two categories, i.e., Category A and Category B, and we have a new data point  $x_1$ , so this data point will lie in which of these categories[12]. To solve this type of problem, we need a K-NN algorithm. With the help of K-NN, we can easily identify the category or class of a particular dataset(as shown in Fig.2). Consider the below diagram:

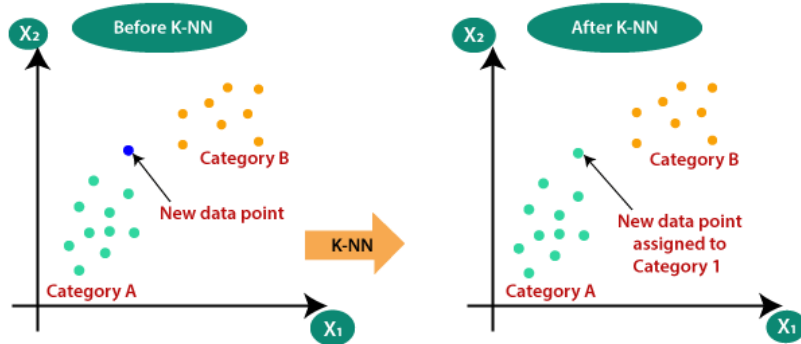


Fig.2. Category A and B differentiation

**4. XGBOOST Method:**

XGBoost is a decision-tree-based ensemble Machine Learning algorithm that uses a gradient boosting framework. In prediction problems involving unstructured data (images, text, etc.) artificial neural networks tend to outperform all other algorithms or frameworks. However, when it comes to small-to-medium structured/tabular data, decision tree based algorithms are considered best-in-class right now. Please see the chart below for the evolution of tree-based algorithms over the years[13].

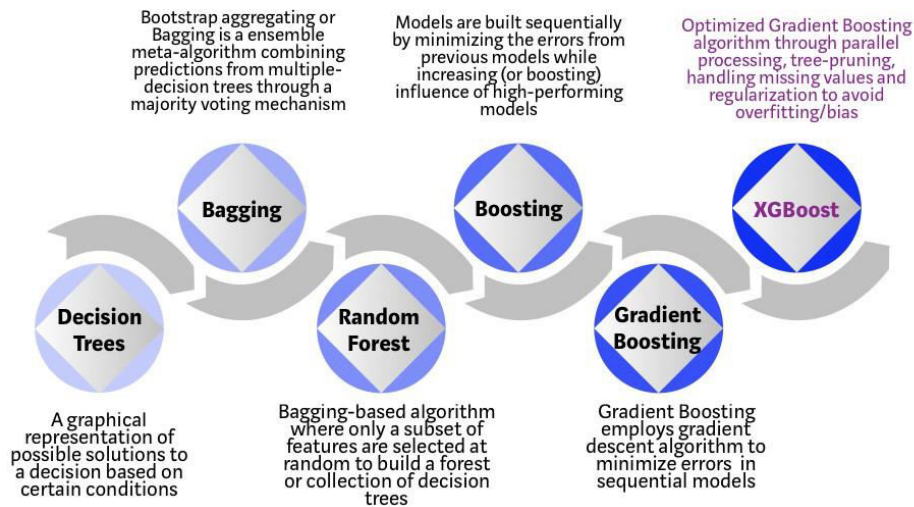


Fig.3. Evolution of XGBoost Algorithm from Decision Trees

XGBoost algorithm was developed as a research project at the University of Washington. Tianqi Chen and Carlos Guestrin presented their paper at SIGKDD Conference in 2016 and caught the Machine Learning world by fire. Since its introduction, this algorithm has not only been credited with winning numerous Kaggle competitions but also for being the driving force under the hood for several cutting-edge industry applications.

**V. CONCLUSION**

All the three different types of PD data patterns used in practice have been discussed. A Comprehensive picture of the methods used by other researches for recognition of PD type and their improvised techniques used by this research already presented in papers [14-20] and some more yet to publish has been discussed in this paper.

A Comprehensive picture of the state-of-the-art knowledge on the feature extraction (statistical method) and classification methods (GNB,KNN and Xgboost) of PD signals has been presented in this paper. Various algorithms tools proposed for both evaluation of feature parameters and their classification for the various data patterns have been

described in sufficient detail so as to enable the fresh researchers in this area to grasp the essence of the present day technology[1].

#### REFERENCES

- [1] N.C. Sahoo, M. M. A. Salama, R. Bartnikas (2005), "Trends in Partial Discharge Pattern Classification: A Survey", IEEE Transactions on Dielectrics and Electrical Insulation, Vol. 12, No. 2; April 2005.
- [2] Wen-Yeou Chang and Hong-Tzer Yang, "Application of Self organizing map approach to partial discharge pattern recognition of Cast resin current transformers, WSEAS transactions on computer research.
- [3] E. Gulski and A. Krivda(1993), "Neural Networks as a Tool for Recognition of Partial Discharges", IEEE Transactions on Electrical Insulation, Vol. 28 No.8, December 1993.
- [4] Burges C., "A tutorial on support vector machines for pattern recognition", In "Data Mining and Knowledge Discovery". Kluwer Academic Publishers, Boston, 1998, (Volume 2).
- [5] L. Hao and P. L. Lewin(2010), "Partial Discharge Source Discrimination using a Support Vector Machine, IEEE Transactions on Dielectrics and Electrical Insulation Vol. 17, No. 1; February 2010.
- [6] A. Contin, A. Cavallini, G. C. Montanari, C. Hudon, M. Belec, D. N. Nguyen, "Searching for Indexes Suitable for Rotating Machines Diagnosis", IEEE Intl. Symposium on Electrical Insulation, pp.101-105, 2006.
- [7] M. Mansor, P. S. Ghosh, A. B. Abdul Ghani, "PD Patterns Recognition in XLPE Cable Under various thermal conditions using Statistical Technique", Intl. Journal Research and Reviews in Applied Sciences(IJRRAS), Vol. 5, Issue 2, pp. 11-181, 2010.
- [8] A. Krivda, —Automated Recognition of Partial Discharges, IEEE TDEI, Vol. 2, No. 5, pp. 796-821, 1995.
- [9] A. Krivda, E. Gulski, L. Satish, W. S. Zaengl, —The Use of Fractal Features for Recognition of 3-D Discharge Patterns, IEEE TDEI, Vol. 2, pp. 889-892, 1995.
- [10] A. Contin, E. Gulski, M. Cacciari, G. C. Montanari, —Applications of the Weibull Function to Partial Discharge Data coming from Different sources Typologies, Electrical Insulation and Dielectric Phenomena, pp. 335- 338, 1995.
- [11] M. Cacciari, A. Contin, G. Rabach, G. C. Montanari, —Diagnosis of Practical Insulation Systems by Partial Discharge Measurements in the Presence of Multi-discharge Phenomenal, Intl. Conference on Electrical Insulation and Dielectric Phenomena, pp. 414- 419, 1993.
- [12] B. Freisleben, M. Hoof, R. Patsch, —Using Counter propagation Neural Networks for Partial Discharge Diagnosis, Intl. Journal of Neural Computing and Applications, Volume 7, Issue 4, pp. 318-333, 1998.
- [13] R. Candela, G. Mirelli, R. Schifani, —Partial Discharge Recognition by Means of Statistical and Fractal Parameters and a Neural Network, IEEE TDEI, Vol. 7, No. 1, pp. 87-94, 2000.
- [14] Namrata Bhosale, Priyanka Kothoke Amol Deshpande, Dr. Alice Cheeran, "Analysis of Partial Discharge using Phase-Resolved( $\phi$ -q) and ( $\phi$ -n) Statistical Techniques", International Journal of Engineering Research and Technology, Vol. 2 (05), 2013,ISSN2278-0181.
- [15] Priyanka Kothoke, Namrata Bhosale, Amol Deshpande, Dr. Alice Cheeran, "Analysis of Partial Discharge using Phase-Resolved (n-q) Statistical Techniques", International Journal of Engineering Research and Applications.
- [16] Yogesh R. Chaudhari, Namrata R. Bhosale, Priyanka M. Kothoke, "Composite Analysis of Phase Resolved PD Patterns using Statistical Techniques", International Journal of Modern Engineering Research (IJMER) Vol. 3, Issue. 4, pp-1947-1957, 2013.
- [17] Priyanka Kothoke, Anupama Deshpande, Yogesh Chaudhari, "A Literature Review of methods used for Partial Discharge Type Detection and its Improved Techniques", International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET), Vol. 8, Issue 7, July 2019.
- [18] Priyanka Kothoke, Dr. Anupama Deshpande, Yogesh Chaudhari, "Investigation and Determination of PD source utilizing SVR and RF", International Conference on Inventive Computation Technologies (ICICT-2020), inclusion into IEEE Xplore and Scopus
- [19] Priyanka Kothoke, Dr. Anupama Deshpande, Yogesh Chaudhari, "Examination and Determination of Partial Discharge Source utilizing SOM and BPM Techniques of ANN", International Journal of Advanced Science and Technology, Vol. 29, No. 8s, (2020), pp. 3299-3306.
- [20] Priyanka Kothoke, Dr. Anupama Deshpande, Yogesh Chaudhari, "Analysis and Determination of Partial Discharge Type using Statistical Techniques and Back Propagation Method of Artificial Neural Network for Phase Resolved Data;", International Journal of Engineering Research and Technology(IJERT),Vol. 08, Issue 08, August 2019.



**Impact Factor:  
7.569**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details