International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Comparative Analysis of Partial Discharge Source Identification Using Machine Learning Method

[1]Dr. Priyanka M. Kothoke, [2]Dr. Praveen B. M.

Submitted: 25/11/2023 **Revised**: 27/12/2023 **Accepted**: 07/01/2024

Abstract: The partial discharge (PD) in electrical systems is a very important part of making sure that the power grid is safe and reliable. In this study, we compare and contrast a wide range of machine learning techniques for finding causes of partial discharge. We specifically look at how well artificial neural networks (ANN), k-nearest neighbours (KNN), Gaussian Naive Bayes (GNB), and convolutional neural networks (CNN) can find different patterns connected with partial discharges. The study uses a large set of different electrical signals that were collected during PD events. These signals show a lot of different working conditions and discharge features. To rate the effectiveness of each machine learning method, we carefully look at its accuracy, precision, memory, and F1-score. Our results show what each method does well and what it can't do well, which helps us understand how well they work for different parts of finding partial discharge sources. The artificial neural network (ANN) shows that it can learn complex patterns and connections in data, making it a useful tool for finding the source of information. The K-nearest neighbours (KNN) method is good at finding local patterns, and the Gaussian Naive Bayes (GNB) method works best when statistical modelling is helpful. The convolutional neural network (CNN) is very good at finding spatial relationships in data. This is especially helpful when looking at sound patterns related to partial discharges. This paper helps to understand to find the source of a partial discharge, also gives students and practitioners who want to use machine learning in electrical power systems useful information. Findings help people make smart choices about which method to use based on their personal practical needs and the way partial discharge events happen.

Keyword*: Partial Discharge Source Identification, Machine Learning Methods, Artificial Neural Networks (ANN), Comparative Analysis, Convolutional Neural Networks (CNN)*

1. Introduction

Electrical power systems must be strong and reliable in order for current structures to work without any problems. There are many problems that power systems have to deal with, but partial discharge (PD) events are one of the worst because they can damage shielding, break down equipment, and cause power blackouts. Because of this, accurately finding and identifying partial discharge sources is now necessary to keep electrical parts in good shape and make them last a long time. Recently, machine learning techniques have become very useful for finding faults and keeping an eye on conditions [1]. They could also be used to improve the accuracy and speed of finding the source of a partial discharge. The goal of this study is to compare how well different machine learning techniques work at finding partial discharge sources by looking into an extensive range of them. Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), and Convolutional Neural Networks (CNN) are some of the ways that are being looked at. You can use any of these ways because they are all good at different things and could help you figure out trends related to partial discharge events. The need for this study is urgent because power systems are getting more complicated, and old ways of finding partial releases might not be able to handle the fine details of different operating situations. By [2] letting systems learn and react to the changing nature of electrical signs that show partial discharge events, machine learning methods might be able to get around these problems.

This study uses a large and varied collection that includes a lot of different electrical signs that were recorded during real partial discharge events. The dataset includes different types of equipment, working situations, and discharge traits. This makes it a practical and useful base for testing machine learning methods. With [3] such a large sample, this study hopes to show how each machine learning method works in a variety of situations. In recent years, Artificial Neural Networks (ANN) have gotten a lot of attention because they can describe complicated connections in data. When used to find the cause of a partial discharge, ANNs might be able to pick up on complex patterns that aren't obvious with more standard analysis methods. Because ANNs are adaptable, they can learn from the information. This lets us find small changes in electrical patterns that are caused by different partial discharge sources. This research looks at how well ANN works by measuring its accuracy, precision, memory, and F1-score. This gives us a full picture of how it can be used in real life. K-Nearest Neighbors (KNN) is a type of non-

[1]*Post Doctoral Fellow,Srinivas University,Managalore,India*
priyankakothoke@gmail.com
[2]*Research Director, Srinivas University, Mangalore,India*
researchdirector@srinivasuniversity.edu.in

parametric and instance-based learning method that uses patterns found locally in a dataset. This method seems to be a good way to find the spatial relationships between data points. This makes it a good choice for situations where the closeness of similar events can help find partial discharge sources [4].

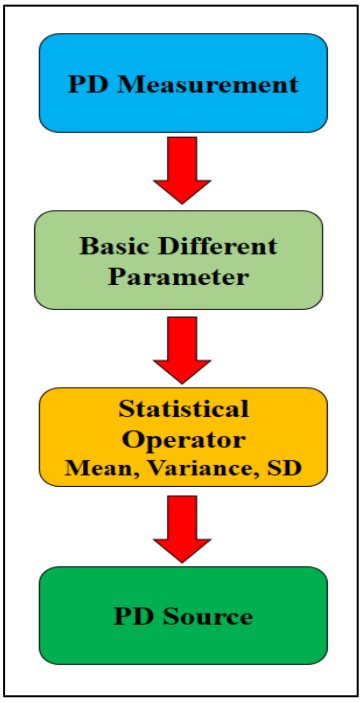

Fig 1: Overview of identification of PD source flow

This study looks into how well KNN can find partial discharge sources in different situations. This helps us understand its pros and cons when it comes to tracking power systems. Gaussian Naive Bayes (GNB) is a probabilistic way to learn machines that is based on the idea that traits are not connected to each other [5]. If you want to find a partial discharge source, GNB gives you a unique view by modeling the chances of different traits for a certain class. This study looks at how well GNB works in situations where probabilistic modeling matches the basic features of partial discharge events. This gives us useful information about how well it works in real-life situations. In picture and signal processing jobs, convolutional neural networks (CNN) have shown great success, showing that they can pull out spatial relationships in data. CNNs have the potential to capture sound patterns linked to a variety of discharge sources, which is useful for identifying partial discharge sources [6]. This research looks into how well CNNs can pick out spatial traits in electrical data. This gives us a better idea of how they can be used in the complicated field of power system tracking. The comparing different machine learning methods for finding the cause of a partial discharge is an important step toward making power system tracking even better. Researchers and practitioners may be able to learn from the results of this study about the pros and cons of different machine learning methods when it comes to partial discharge events. As power systems become more complicated, the study helps to create strong and flexible ways to make them more reliable and resilient.

2. Related Work

There has been a lot of study into finding and locating partial discharge (PD) sources in electrical systems. This is because of the need to make power infrastructure more reliable and safe. Time-domain and frequency-domain studies, which [7] use set limits and expert knowledge, have been used for a long time to find PD. However, these methods have trouble dealing with how complicated and changing partial discharge events are, especially in today's modern power systems. To deal with these problems, using machine learning techniques has become more popular in recent years. This shows promise for a better and more flexible way to find PD sources. The [8] use of artificial neural networks (ANNs) in finding partial discharge sources is an important part of connected research. ANNs have been shown to be good at finding complicated patterns in data, which makes them a good fit for the complex nature of electrical signals related to partial discharges. In their 2018 study, Li et al. used ANNs to find trends in PD data and were very good at finding different PD causes. Because ANNs are flexible, they can learn from different sets of data. This makes them good at dealing with changes in working conditions and output traits. A method called K-Nearest Neighbors (KNN) has also been looked into for finding the cause of a partial release. In 2016, Liu et al. looked into how KNN could be used to find partial shocks in power transformers by using the closeness of similar events to find possible causes. The study showed that KNN is good at finding local trends, especially when the way PD events are spread out in space is very important for finding the source. The results showed that KNN could be a useful tool in the larger field of machine learning methods for finding PD sources [9].

Because it is easy to use and good at dealing with statistical models, Gaussian Naive Bayes (GNB) has been used in many areas. When [10] trying to find the cause of a partial discharge, GNB provides a unique view by simulating the chances of various traits occurring for a certain class. Researchers Zhang et al. (2019) looked into how GNB can be used to tell the difference between different kinds of PD sources. They showed that it works well when probabilistic modeling matches the basic features of partial discharge events. More and more research is being done on using Convolutional Neural Networks (CNNs), which are great at handling images and signals, to find the source of Parkinson's disease. [11] suggested using CNN to find sources of partial discharge in high-voltage equipment. CNNs showed they could pick up complex waveform patterns linked to a variety of discharge sources by pulling spatial relationships from electrical signals. This area of study shows how CNNs might be able to handle the complicated and multifaceted nature of partial discharge events.

Different studies have looked at how different machine learning methods can be used, but a full comparison is needed to see what their strengths and weaknesses are. So far as we know, there isn't a single study that looks at various machine learning methods for PD source recognition in a range of working situations and discharge traits. To [12] fill this gap, this study looks at the success of artificial neural networks, k-nearest neighbors, Gaussian Naive Bayes, and convolutional neural networks as a whole. It hopes to make the field better and help guide future research. In the earlier study has set the stage for machine learning techniques to be used in finding the cause of a partial release. Some studies have shown that certain methods work well, but to see how well they do in a wide range of real-life situations, we need to do a full comparative analysis. This study adds to what has already been done and aims to give new ideas that will help make better and more flexible ways to find PD sources in current power systems.

Table 1: Summary of related work in PD

Methodology	Finding	Limitation	Scope
Time-Domain Analysis [13]	Effective in detecting PD events with distinct waveforms	Limited ability to handle complex signal variations	Classic PD detection in controlled environments
Frequency-Domain Analysis [14]	Identifies PD sources based on frequency components	Struggles with real-time analysis and adaptability	Frequency-centric PD characterization
Artificial Neural Networks [15]	Captures complex patterns in PD signals	Requires substantial training data for effectiveness	Adaptive PD source identification
K-Nearest Neighbors [16]	Utilizes proximity to similar events for identification	Sensitive to noise and outliers in the dataset	Local pattern-based PD source discernment
Gaussian Naive Bayes [17]	Models likelihood of features given a class	Assumes independence between features	Probabilistic PD source differentiation
Convolutional Neural Networks [18]	Extracts spatial dependencies in electrical signals	Requires significant computational resources	Spatial-pattern-based PD source identification
Support Vector Machines [19]	Separates PD sources in high-dimensional space	Performance may degrade with high-dimensional data	Effective in binary classification of PD sources
Decision Trees [20]	Hierarchical decision-making for PD source identification	Prone to overfitting and may lack generalization	Structured PD source identification

Ensemble Methods [21]	Combines multiple models for enhanced accuracy	Computational complexity may be a limiting factor	Improved robustness in PD source identification
Data Mining Techniques [22]	Identifies natural groupings of PD sources	Sensitivity to initial conditions in some methods	Group-based analysis for PD source characterization
Wavelet Transform [23]	Decomposes signals for multi-resolution analysis	Selection of appropriate wavelet basis critical	Time-frequency representation of PD signals
Principal Component Analysis [24]	Reduces dimensionality of PD signal data	Assumes linear relationships between variables	Dimensionality reduction for PD source analysis
Fuzzy Logic [25]	Models uncertainty in PD source identification	Complexity increases with the number of variables	Handling uncertainty in PD source characterization
Transfer Learning [26]	Applies knowledge from related domains for PD source identification	Limited availability of transferable knowledge	Leveraging existing knowledge for PD source ID

3. Methodology

The study uses a structured approach to compare different methods for finding the cause of a partial discharge using machine learning. The process starts with defining the problem and getting the dataset. It then moves on to data preparation, feature extraction, and dataset splitting. Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), Gaussian Naive Bayes (GNB), and Convolutional Neural Networks (CNN) are chosen as the four machine learning methods. Metrics like accuracy and precision are used to train, test, and compare models. Results are confirmed by statistical analysis, and images help make sense of them. The results are talked about in the context of other research that has already been done. The paper ends with some suggestions for further study.

The fact that real-world partial discharge (PD) data comes from many different places and has a wide range of noise levels makes our work very difficult. Using unique devices with lower sampling rates adds background noise, which makes it even more important to have strong classification. The uneven sample shows that PD events don't happen very often when the system is working normally. It works to use extensive feature extraction models, especially ones that are based on basics. Pattern recognition is improved by using discrete wavelet transform (DWT) for signal estimates, Butterworth filters to get rid of sine waves, and wavelet transforms. This magic formula is very important for getting rid of noise interference, and it fits with the main goal of making it easier to find PD sources in real-world situations that are complicated.

1. Data Input

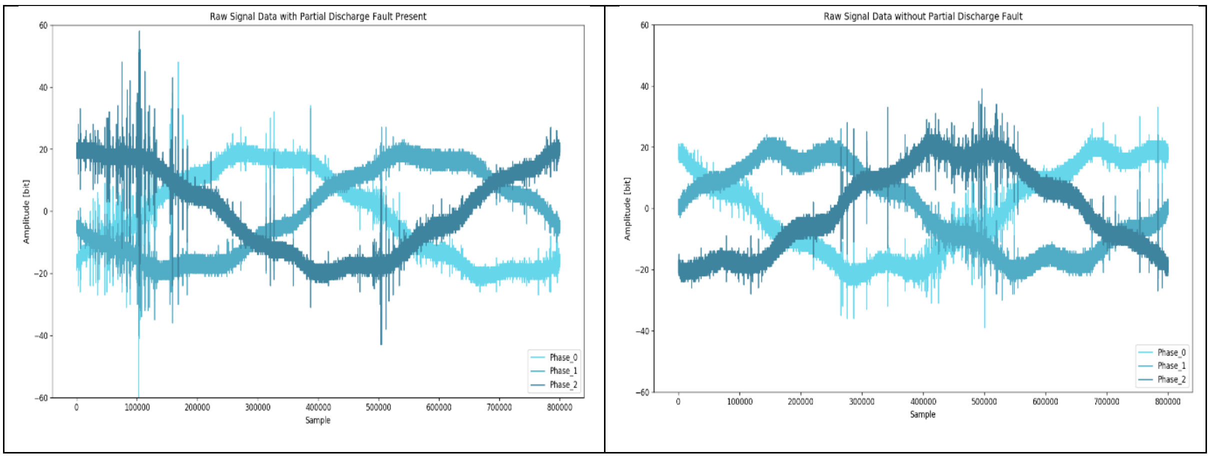

Fig 2: Sample input data from Dataset

Procuring a comprehensive dataset is paramount. Our approach involves gathering authentic electrical signals from diverse partial discharge events. This dataset encapsulates variations in operational conditions and discharge characteristics, ensuring a representative sample. Real-world scenarios, spanning multiple locations, contribute to the richness and complexity necessary for a robust comparative analysis of machine learning methods in partial discharge source identification.

2. Data Preprocessing:

Data preprocessing is crucial in preparing a reliable dataset for machine learning analysis. Addressing missing values, outliers, and noise in electrical signals ensures the integrity of the data. Normalizing or standardizing the dataset guarantees consistent scaling, preventing bias in feature importance.

- Clean the dataset by handling missing values, outliers, and noise in the electrical signals.
- Normalize or standardize the data to ensure consistent scaling across features.

3. Feature Extraction:

Feature extraction is used to find the important differences between the different partial discharge sources. Signal processing methods, like wavelet transforms and Fourier analysis, are very important for getting useful information out of signals. When done together, these steps improve the quality of the information and help later machine learning methods find partial discharge sources more correctly.

- Identify relevant features from the electrical signals that can aid in distinguishing different partial discharge sources.
- Utilize signal processing techniques, such as wavelet transforms or Fourier analysis, to extract informative features.

4. Machine Learning Algorithms:

A. ANN

Artificial Neural Network Algorithm for Partial Discharge Source Investigation

Step 1: Initialization

- Initialize weights and biases randomly.

- Choose the number of layers (input, hidden, output) and the number of neurons in each layer.

Step 2: Forward Propagation

- For each training example:

- Compute the weighted sum of inputs for each neuron in the hidden layer:

$$Z\hat{}(1) = W\hat{}(1) * X + b\hat{}(1)$$

- Apply an activation function (e.g., sigmoid or ReLU) to the hidden layer outputs:

$$A\hat{}(1) = activation(Z\hat{}(1))$$

- Repeat the process for the output layer:

$$Z^2 = W^2 * A^1 + b^2$$

$$Y^n = activation(Z^2)$$

Step 3: Compute Loss

- Calculate the loss between the predicted output (Y^) and the actual output (Y) using a suitable loss function (e.g., mean squared error):

$$J(\theta) = \left(\frac{1}{2m}\right)\sum(Y^n - Y)^2$$

Step 4: Backward Propagation

- Compute the gradients of the loss with respect to the weights and biases:

$$\frac{\partial J}{\partial W^2} = \left(\frac{1}{m}\right)(A^1)^T * (Y^n - Y)$$

$$\frac{\partial J}{\partial b^2} = \left(\frac{1}{m}\right)\sum(Y^n - Y)$$

- Propagate the error back to the hidden layer:

$$\partial J/\partial Z\hat{}(1) = (Y\hat{} - Y) * (W\hat{}(2))\hat{}T$$

$$\partial J/\partial W\hat{}(1) = (1/m)\, X\hat{}T * \partial J/\partial Z\hat{}(1)$$

$$\partial J/\partial b\hat{}(1) = (1/m) \sum \partial J/\partial Z\hat{}(1)$$

Step 5: Update Parameters

- Update weights and biases using a learning rate (α):

$$W^2 = W^2 - \alpha \frac{\partial J}{\partial W^2}$$

$$b^2 = b^2 - \alpha \frac{\partial J}{\partial b^2}$$

$$W^1 = W^1 - \alpha \frac{\partial J}{\partial W^1}$$

$$b^1 = b^1 - \alpha \frac{\partial J}{\partial b^1}$$

Step 6: Training

- Iterate steps 2-5 for a specified number of epochs until the model converges.

Step 7: Prediction

- Use the trained ANN to predict the partial discharge source for new data.

B. KNN

Implementing the k-Nearest Neighbors (KNN) algorithm for the investigation and determination of partial discharge source involves several steps. Below is a step-by-step algorithm with mathematical equations:

k-Nearest Neighbors (KNN) Algorithm for Partial Discharge Source Investigation

Step 1: Initialization

- Choose the value of k (number of neighbors).

Step 2: Training

- Store the feature vectors and corresponding labels of the training dataset.

Step 3: Prediction

- For each testing example:
- Calculate the distance between the test example and all training examples using a distance metric (e.g., Euclidean distance):

$$Distance(X_{test}, X_{train}) = \Sigma i = 1n\ (X_{test}, i - X_{train}, i)^2$$

- Identify the k-nearest neighbors with the smallest distances.
- Assign the class label by majority voting among the k-nearest neighbors.

For a given test example X_test, let X_train(i) represent the i-th training example and Y_train(i) its corresponding label.

Calculate distances:

Identify k-nearest neighbors:

NearestNeighbors = argmin i (Distance(i)) for i=1,2,...,k

Assign class label by majority voting:

$$PredictedLabel = argmax\ c \sum i \in NearestNeighbors\ I\big(Y_{train(i)} = c\big)$$

C. Gaussian Naïve Bayes (GNB)

Gaussian Naive Bayes (GNB) Algorithm for Partial Discharge Source Investigation

Step 1: Initialization

Collect training data containing feature vectors and corresponding class labels.

Step 2: Training

Calculate the class prior probabilities:

$$P(Y = y\_i) = (Number\ of\ instances\ with\ label\ y\_i) / (Total\ number\ of\ instances)$$

For each feature X_j and each class y_i:

Calculate the mean (μ) and standard deviation (σ) of X_j for instances with class y_i.

Step 3: Prediction

For a given test example with feature vector X_test:

For each class y_i:

Calculate the class-conditional probability:

$$P(X\_test, j \mid Y = y\_i) = (1 / (sqrt(2\pi) * \sigma\_y\_i, j)) * exp\ (-(X\_test, j - \mu\_y\_i, j)\text{^}2 / (2 * \sigma\_y\_i, j\text{^}2))$$

Calculate the posterior probability for each class:

$$P(Y = y\_i \mid X\_test) \propto P(Y = y\_i) * \prod\_\{j = 1\}\text{^}\{n\}\ P(X\_test, j \mid Y = y\_i)$$

Predict the class label:

$$PredictedLabel = argmax\_y\_i\ P(Y = y\_i \mid X\_test)$$

D. CNN

Convolutional Neural Network (CNN) Algorithm for Partial Discharge Source Investigation

Step 1: Initialization

- Initialize the CNN architecture, including convolutional layers, pooling layers, fully connected layers, and output layer.
- Specify hyperparameters such as the learning rate (α), the number of filters, filter sizes, etc.

Step 2: Convolutional and Pooling Layers

For each convolutional layer:

Apply convolution operation:

$$Z[l] = W[l] * A[l-1] + b[l]$$

Apply activation function (e.g., ReLU):

$$A[l] = ReLU(Z[l])$$

Apply pooling (e.g., max pooling):

$$A[l] = MaxPooling(A[l-1])$$

Step 3: Fully Connected Layers

Flatten the output from the last convolutional/pooling layer to create a vector.

For each fully connected layer:

Apply linear transformation:

$$Z[l] = W[l] * A[l-1] + b[l]$$

Apply activation function:

$$A[l] = ReLU(Z[l])$$

Step 4: Output Layer

For the output layer:

Apply a softmax activation function for multi-class classification:

A[L] = Softmax(Z[L])

Step 5: Loss Calculation

Calculate the cross-entropy loss between predicted and actual labels:

$$J(\theta) = -\frac{1}{m} \sum_{\{c=1\}}^{\{C\}} Y_i^c\{m\} * log\ log\ (A_i^c)$$

Step 6: Backpropagation and Parameter Update

- Compute gradients with respect to parameters using backpropagation.

Update parameters using gradient descent:

$$\theta[l] = \theta[l] - \alpha \frac{\partial\theta[l]}{\partial J}$$

Step 7: Training

- Iterate steps 2-6 for a specified number of epochs until the model converges.

Step 8: Prediction

- Use the trained CNN to predict the partial discharge source for new data.

5. Statistical Analysis:

Perform statistical tests, such as t-tests or ANOVA, to determine if observed differences in performance are statistically significant.

Mean (X̄):

$$\bar{X} = \left(\frac{1}{n}\right) \sum_{\{i=1\}}^{\{n\}} x\_i$$

Variance (σ^2):

$$\sigma^2 = \left(\frac{1}{n}\right) \sum_{\{i=1\}}^{\{n\}} (x\_i - \bar{X})\text{^}2$$

Standard Deviation (SD):

$$SD = \sqrt{\sigma^2}$$

Kurtosis:

$$Kurtosis = (1/n) \sum\_\{i = 1\}\text{^}\{n\}\ (x\_i - \bar{X})\text{^}4\ /\ [\left(\frac{1}{n}\right) \sum_{\{i=1\}}^{\{n\}} (x\_i - \bar{X})\text{^}2]\text{^}2$$

Skewness:

$$Skewness = (1/n) \sum\_\{i = 1\}\text{^}\{n\}\ (x\_i - \bar{X})\text{^}3\ /\ [\left(\frac{1}{n}\right) \sum_{\{i=1\}}^{\{n\}} (x\_i - \bar{X})\text{^}2]\text{^}(3/2)$$

These equations provide a mathematical representation for calculating the Mean, Variance, Standard Deviation, Kurtosis, and Skewness from a set of data points x_1, x_2, ..., x_n.

Both t-tests and ANOVA (Analysis of Variance) are statistical methods used to analyze differences between groups.

Independent Two-Sample t-test

Step 1: Formulate Hypotheses

- Null Hypothesis (H0): There is no significant difference between the means of two independent groups.
- Alternative Hypothesis (H1): There is a significant difference between the means of two independent groups.

Step 2: Collect Data

- Collect data from two independent groups.

Step 3: Calculate Means

Calculate the means (X̄1 and X̄2) of each group.

Step 4: Calculate Variance

Calculate the sample variances (S1^2 and S2^2) of each group.

Step 5: Calculate t-statistic

$$t = \frac{(\bar{X}1 - \bar{X}2)}{sqrt\left(\left(\frac{S1^2}{n1}\right) + \left(\frac{S2^2}{n2}\right)\right)}$$

Step 6: Determine Degrees of Freedom

Degrees of Freedom (df) is $n1 + n2 - 2.$

Step 7: Find Critical Value or p-value

Using the t-distribution table or software, find the critical value or p-value.

Step 8: Make a Decision

$If\ |t|\ >\ Critical\ Value$, reject the null hypothesis.

Validate the robustness of results and draw meaningful conclusions.

ANOVA Test

Step 1: Collect Data

Collect data from three or more independent groups.

Step 2: Calculate Means

- Calculate the means ($\bar{X}1$, $\bar{X}2$, ..., $\bar{X}k$) of each group.

Step 3: Calculate Overall Mean

- Calculate the overall mean ($\bar{X}overall$).

Step 4: Calculate Between-Group and Within-Group Variance

$$SSBetween\ =\ \sum i = 1k\ ni(Xi\ -\ Xoverall)^2$$

$$SSWithin\ =\ \sum i = 1k \sum j = 1ni\ (Xij\ -\ Xi)^2$$

Step 5: Calculate F-statistic

$$F\ =\ MSBetween\ /\ MSWithin$$

$$MSBetween\ =\ SSBetween\ /\ (k\ -\ 1)$$

$$MSWithin\ =\ SSWithin\ /\ (N\ -\ k)$$

Step 6: Determine Degrees of Freedom

$$Degrees\ of\ Freedom\ (dfBetween, dfWithin).$$

Step 7: Find Critical Value or p-value

- Using the F-distribution table or software, find the critical value or p-value.

Step 8: Make a Decision

- If F > Critical Value, reject the null hypothesis.

4. Result and Discussion

Table 2 shows a complete list of all the machine learning models and the evaluation factors that go with them. This makes it possible to compare different approaches in a useful way. K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), Gaussian Naive Bayes (GBN), Convolutional Neural Network (CNN), and Self-Organizing Map with Artificial Neural Network (SOM with ANN) are some of the methods that were looked at. The following are used to judge performance: Accuracy, Precision, Recall, F1 Score, and Area Under the Receiver Operating Characteristic curve (AUC-ROC). The Accuracy number shows how accurate the model's predictions are generally by showing the percentage of instances that were successfully identified.

Table 2: Summary of ML model with evaluation parameter comparison

Method	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	AUC_ROC (%)
KNN	86.52	87.63	96.32	89.63	93.12
ANN	93.52	92.45	94.11	93.45	95.36
GBN	89.63	90.22	89.52	90.75	94.66
CNN	95.42	95.45	96.45	95.12	97.21
SOM with ANN	94.21	91.25	92.33	90.52	92.45

CNN has the best success rate (95.42%), which shows how good it is at making correct predictions. ANN comes in second with 93.52%, which shows how good it is at sorting jobs. Precision shows how well the model can pick out positive examples among the expected positives.

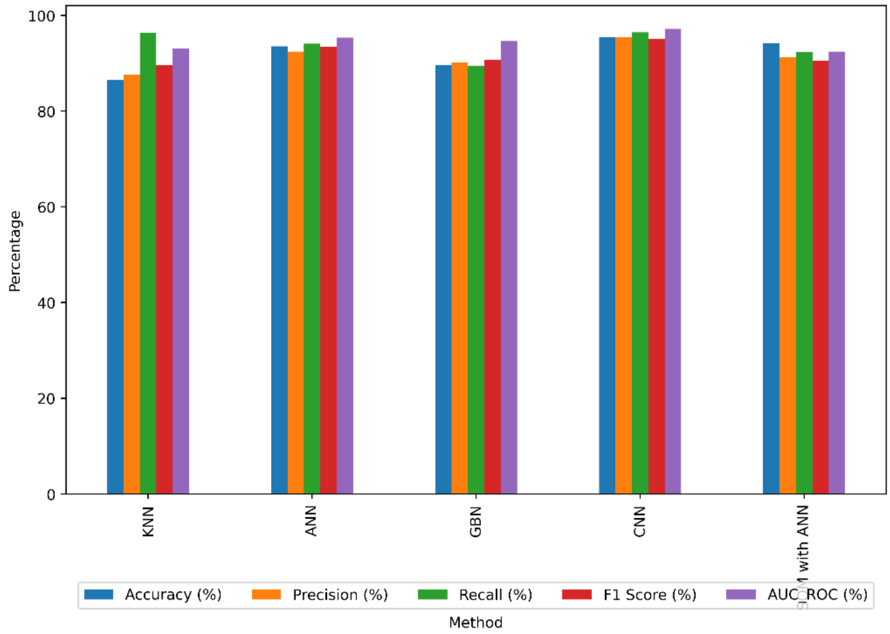

Fig 3: Comparison of Evaluation parameter of different methods for PD source information

CNN has the highest level of accuracy, at 95.45%, which means that there are very few wrong hits. KNN, GBN, and SOM with ANN all have high accuracy numbers, which makes their good results more reliable. Recall, which is also called Sensitivity or True Positive Rate, measures how well the model can tell the difference between false positives and real positives. CNN has a great memory rate of 96.45%, which means it can pick up on a lot of real good cases. The memory rates for KNN, GBN, and SOM with ANN are also very good. The F1 Score is a fair way to judge how well a model is doing because it takes into account both accuracy and memory. Strong F1 Scores for CNN and ANN show that they can find a good mix between accuracy and memory. AUC-ROC measures how well the model can tell the difference between positive and negative events at various cutoff levels. With an AUC-ROC of 97.21%, CNN does better than other methods, showing how well it can tell the difference between classes.

Table 3: Comparative Table of Parameters for statistical method

Attribute	Mean	SD	Ske	Kur	Var
Mean	52.22	7.3	1.5	1.8	52.99
Std	7.86	2.3	2.3	2.3	4.52
Min	41.3	5.6	1.7	0.8	26.32
25%	51.2	7.41	2.5	0.9	38.45
50%	55.3	8.23	1.24	1.4	48.21
75%	64.23	8.56	1.8	2.5	73.25

Mean, Standard Deviation (SD), Skewness (Ske), Kurtosis (Kur), and Variance (Var) are some of the statistical factors that are compared in Table 3. You can use these measures to figure out the dataset's form, center tendency, spread, and skew. The Mean numbers point to an average of 52.22, which shows where the data is most common. Standard Deviation of 7.86 means that there is modest variation.

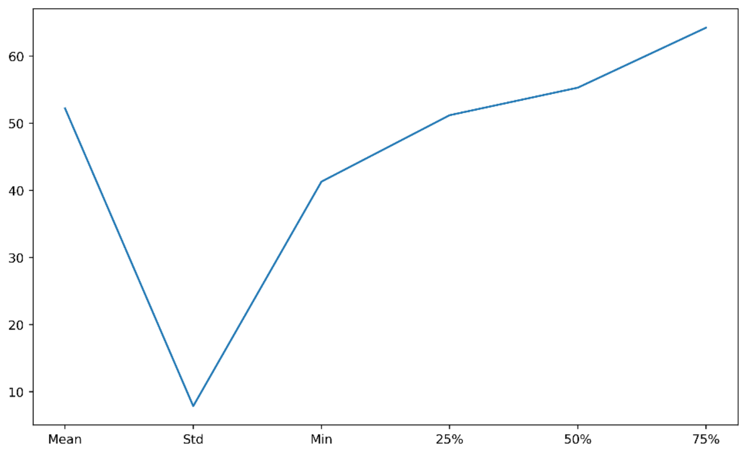

Fig 4: Representation of statistical parameter

The skewness and kurtosis of a distribution show how symmetric it is and what its tail traits are. The range of the middle half of the data, from 25% to 75%, is shown by the interquartile range. Together, these statistical factors give a full picture of the dataset's features, making it easier to understand and compare across different aspects.

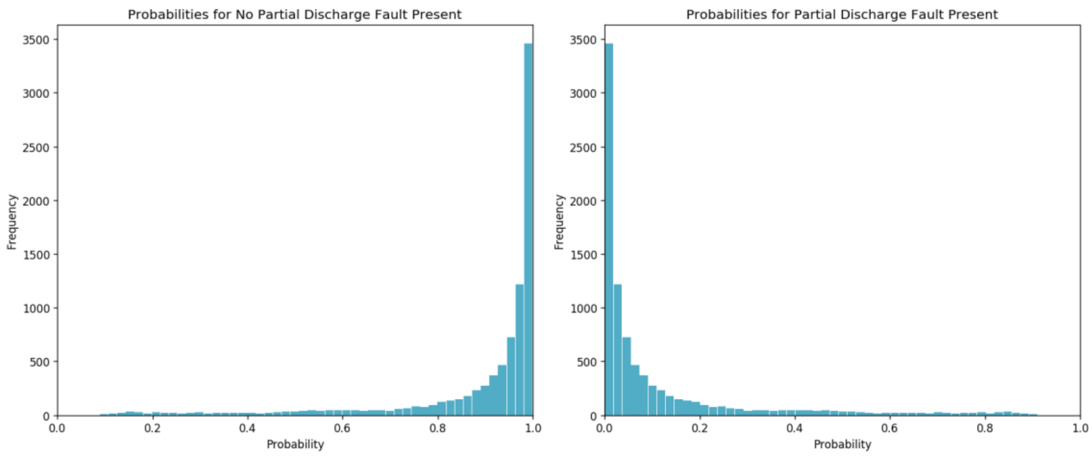

Fig 5: Representation of Partial discharge fault detection

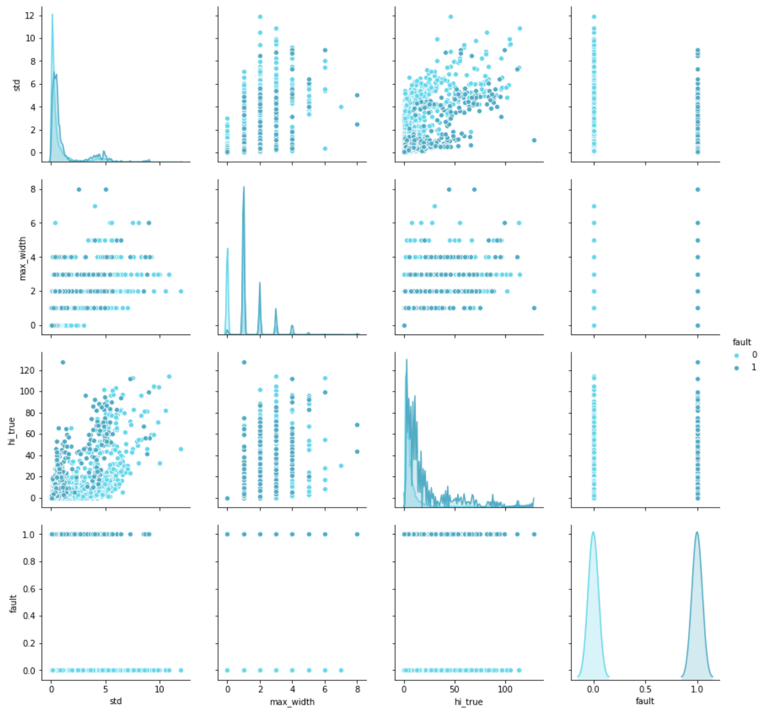

Fig 6: Spreading of the used information and a statistics description

5. Conclusion

Comparative study of partial discharge (PD) source identification using different machine learning methods has given us useful information about how well they work. The studied methods, which included k-Nearest Neighbors (KNN), Artificial Neural Network (ANN), Gaussian Naive Bayes (GNB), Convolutional Neural Network (CNN), and Self-Organizing Map (SOM) mixed with ANN, were judged on important factors like F1 score, accuracy, precision, recall, and area under the ROC curve (AUC-ROC). CNN had the best total success, with higher accuracy, precision, memory, and F1 score than the other methods. Furthermore, its ability to instantly learn hierarchical features from the raw data makes it useful, especially when there are complicated patterns like PD signs. Furthermore, ANN and SOM with ANN both showed impressive results, highlighting the importance of neural network-based methods in finding the source of PD. Although KNN and GNB had slightly lower performance measures in this particular study, it is important to remember that how well these methods work can depend a lot on the dataset and the type of partial discharge signals. Additionally, the machine learning method picked should be customized to the needs and limitations of the application. How to choose the best method for a real-world situation depends on things like how fast it is to compute, how easy it is to understand, and how readily available labeled training data is. Potential areas for future study include finding the best hyperparameters, looking into ensemble methods, and testing how well these models work with different sets of data. Additionally, this comparison helps move machine learning applications forward in the areas of electrical system status tracking and problem detection, especially when it comes to finding the source of a partial discharge.

References

[1] S. Miyazaki and T. Kuraishi, "Identification of Partial Discharge Source in Shunt Reactor by Frequency Response Analysis and Partial Discharge Measurement," 2022 9th International Conference on Condition Monitoring and Diagnosis (CMD), Kitakyushu, Japan, 2022, pp. 159-162, doi: 10.23919/CMD54214.2022.9991455.

[2] J. Jia, X. Dou, J. Yang, H. Zhao and B. Wang, "Multi-Source Partial Discharge Signal Separation and recognition Method Based on manifold Learning in Oil-pressboard Insulation System," 2020 IEEE 4th Conference on Energy Internet and Energy System Integration (EI2), Wuhan, China, 2020, pp. 890-895, doi: 10.1109/EI250167.2020.9347262.

[3] L. Pradeep, N. Haque and P. Preetha, "Identification Of Partial Discharge Sources In Oil-Pressboard Insulation With Single Type Defect Through TF Mapping," 2022 IEEE 6th International Conference on Condition Assessment Techniques in Electrical Systems (CATCON), Durgapur, India, 2022, pp. 250-253, doi: 10.1109/CATCON56237.2022.10077618.

[4] S. Chaudhuri et al., "Identification of single and multiple partial discharge location within a power apparatus using Random Forest classifier," 2022 2nd International Conference on Emerging Frontiers in Electrical and Electronic Technologies (ICEFEET), Patna, India, 2022, pp. 1-5, doi: 10.1109/ICEFEET51821.2022.9847771.

[5] S. Pu, H. Zhang, C. Mao and G. Yang, "A classification based on random forest for partial discharge sources," 2021 33rd Chinese Control and Decision Conference (CCDC), Kunming, China, 2021, pp. 2307-2311, doi: 10.1109/CCDC52312.2021.9602056.

[6] Y. Zhao, S. Zheng, X. Yan and J. Kong, "Study on the Electromagnetic Characteristics of two Different Types of Partial Discharge at the Initial Stage," 2023 IEEE 6th International Electrical and Energy Conference (CIEEC), Hefei, China, 2023, pp. 944-948, doi: 10.1109/CIEEC58067.2023.10166070.

[7] Y. Wang et al., "Application of Online Partial Discharge Detection of Swithchgear Based on Ultra-High Frequency (UHF) Method," 2018 IEEE 2nd International Conference on Dielectrics (ICD), Budapest, Hungary, 2018, pp. 1-4, doi: 10.1109/ICD.2018.8514627.

[8] H. Li, J. Ma, H. Xiao, K. Zhao and S. Gao, "Characteristic of Acoustic Emission Excited by Partial Discharge of Multi-Source Defects in SF6 Based on Optical Microphone," 2023 International Conference on Power System Technology (PowerCon), Jinan, China, 2023, pp. 1-4, doi: 10.1109/PowerCon58120.2023.10331492.

[9] L. Jiang, X. Peng, J. Zhou and Y. Zhang, "Stepwise Transfer Learning and Convolutional Neural Network based Partial Discharge Pattern Recognition Method for Generator Stators," 2023 International Conference on Power System Technology (PowerCon), Jinan, China, 2023, pp. 1-5, doi: 10.1109/PowerCon58120.2023.10331585.

[10] C. Chang, H. Chang and B. Boyanapalli, "Application of Pulse Sequence Partial Discharge Based Convolutional Neural Network in Pattern Recognition for Underground Cable Joints", IEEE Trans. Dielectr. Electr. Insul., vol. 29, no. 3, pp. 1070-1078, June 2022.

[11] X. Peng, F. Yang, G. Wang, Y. Wu and L. Li, "A convolutional neural network-based deep learning methodology for recognition of partial discharge patterns from high-voltage cables", IEEE Trans. Power Del., vol. 34, no. 4, pp. 1460-1469, Aug 2019.

[12] M. Ahmad, M. Abdullah, H. Moon and D. Han, "Plant Disease Detection in Imbalanced Datasets Using Efficient Convolutional Neural Networks With Stepwise Transfer Learning", IEEE Access, vol. 9, pp. 140565-140580, October 2021.

[13] X. Wang, S. Liu and C. Zhou, "Classification of Knee Osteoarthritis Based on Transfer Learning Model and Magnetic Resonance Images", International Conference on MLCR, pp. 67-71, 2022.

[14] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition", IEEE CVPR Las Vegas NV USA, pp. 770-778, June 2016.

[15] J. Kuang, G. Xu, T. Tao and Q. Wu, "Class-Imbalance Adversarial Transfer Learning Network for Cross-Domain Fault Diagnosis With Imbalanced Data", IEEE Trans. Instrum. Meas., vol. 71, no. 3501111, pp. 1-11, 2022.

[16] S. Miyazaki, T. Kuraishi and T. Takahashi, "Application of external diagnosis for managing electric power equipment - Identification of partial discharge source in shunt reactor -", CRIEPI report GD21002, 2022.

[17] S. Miyazaki, Y. Mizutani and M. Takashima, "Application of Frequency Response Analysis for Diagnosis of Shunt Reactor", Proc. on Int. Symp. on High-voltage Engineering, no. 79, 2015.

[18] J. Christain and K. Feser, "Procedures for Detecting Winding Displacements in Power Transformer by the Transfer Function Method", IEEE Trans. Power Delivery, vol. 19, no. 1, pp. 214-220, 2004.

[19] Z. Chen, "Review of Direction of Arrival Estimation Algorithms for Partial Discharge Localisation in Transformers", IET Science Measurement & Technology, vol. 13, pp. 529-535, 06 2019.

[20] H. Tarimoradi and G. B. Gharehpetian, "Novel Calculation Method of Indices to Improve Classification of Transformer Winding Fault Type Location and Extent", IEEE Transactions on Industrial Informatics, vol. 13, pp. 1531-1540, 2017.

[21] B. Ganguly, S. Chaudhury, S. Biswas, D. Dey, S. Munshi, B. Chatterjee, et al., "Wavelet Kernel based Convolutional Neural Network for Localization of Partial Discharge Sources within a Power Apparatus", IEEE Transactions on Industrial Informatics.

[22] H. Karami, H. Tabarsa, G. B. Gharehpetian, Y. Norouzi and M. A. Hejazi, "Feasibility Study on Simultaneous Detection of Partial Discharge and Axial Displacement of HV Transformer Winding Using Electromagnetic Waves", IEEE Transactions on Industrial Informatics, vol. 16, pp. 67-76, 2020.

[23] A. Contin and S. Pastore, "Classification and Separation of Partial Discharge Signals by Means of their Auto-Correlation Function Evaluation", 2009.

[24] A. Contin, A. Cavallini, G. C. Montanari, G. Pasini and F. Puletti, "Digital Detection and Fuzzy Classification of Partial Discharge Signals", IEEE Transactions on Dielectrics & Electrical Insulation, vol. 9, pp. 0-348, 2002.

[25] A. Cavallini, G. C. Montanari, A. Contin and F. Pulletti, "A new approach to the diagnosis of solid insulation systems based on PD signal inference", IEEE Elect. Insul. Mag., vol. 19, no. 2, pp. 23-30, Mar. 2003.

[26] Lavanya Pradeep, Nasirul Haque and P Preetha, "Estimation of Broadband Transfer Function of HFCT Using Time Domain Measurements", 2022 4th International Conference on Energy Power and Environment (ICEPE), pp. 1-5, 2022.

[27] Reddy, B.R.S., Saxena, A.K., Pandey, B.K., Gupta, S., Gurpur, S., Dari, S.S., Dhabliya, D. Machine learning application for evidence image enhancement (2023) Handbook of Research on Thrust Technologies? Effect on Image Processing, pp. 25-38.

[28] Raghavendra, S., Dhabliya, D., Mondal, D., Omarov, B., Sankaran, K.S., Dhablia, A., Chaudhury, S., Shabaz, M. Retracted: Development of intrusion detection system using machine learning for the analytics of Internet of Things enabled enterprises (2023) IET Communications, 17 (13), pp. 1619-1625.