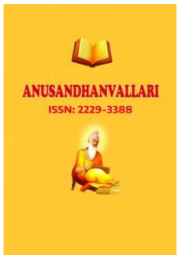

Anusandhanvallari
Vol 2025, No.1
November 2025
ISSN 2229-3388

Comparative Investigation and Determination of Partial Discharge Source Using Gaussian Naive Bayes (GNB)And K- Nearest Neighbour Method

[1]Priyanka Kothoke, [2]Kajol Chaudhari, [3]Anil Kale

[1]Post Doctoral Research Scholar, Apex Professional University,Pasighat, Arunachal Pradesh, India. priyankakothoke250390@gmail.com

[2]Smt.Indira Gandhi College of Engineering, Navi Mumbai, Maharashtra, India. kajolchaudhari20@gmail.com

[3]MGM's college of Engineering and Technology,Panvel, Maharashtra, India. anil5474@gmail.com

Abstract

Partial discharge (PD) detection and classification play a vital role in ensuring the reliability and longevity of high-voltage insulation systems. This study presents a comparative investigation into the identification and determination of PD sources using two supervised machine learning techniques—Gaussian Naive Bayes (GNB) and K-Nearest Neighbor (KNN). Experimental PD data were acquired under controlled conditions from different defect models such as surface discharge, internal discharge, and corona discharge. Key statistical and time–frequency features were extracted from the acquired signals to form an efficient feature set for classification. The GNB algorithm, based on probabilistic reasoning and the assumption of feature independence, provides a fast and computationally efficient classification mechanism. In contrast, the KNN method, a non-parametric approach, classifies data based on similarity measures in multidimensional space. The comparative analysis evaluates both models in terms of accuracy, precision, recall, and computational complexity. Experimental results demonstrate that while KNN yields superior accuracy for nonlinear PD patterns, GNB performs efficiently for real-time and low-complexity scenarios. The study highlights the trade-offs between model performance and computational demand, suggesting a hybrid or ensemble approach for future PD source identification systems in high-voltage applications. This work contributes to improving predictive maintenance and insulation health assessment methodologies.

Keywords: Partial discharge (PD), Machine Learning (ML), Gaussian Naive Bayes (GNB) and K-Nearest Neighbor (KNN).

1. Introduction

Importance of Partial Discharge (PD) Detection in High-Voltage Equipment

High-voltage (HV) electrical equipment such as transformers, gas-insulated switchgear (GIS), cables, and rotating machines are critical components of modern power systems. The reliable performance of these assets ensures uninterrupted power transmission and distribution. However, these systems are constantly subjected to electrical, thermal, mechanical, and environmental stresses that can deteriorate insulation materials over time. Among various insulation degradation phenomena, partial discharge (PD) has been recognized as one of the most significant indicators of insulation weakness or defect. PD refers to localized dielectric breakdowns that do not completely bridge the electrodes but occur within voids, gas bubbles, or along the surface of the insulation. Although each discharge may

Anusandhanvallari
Vol 2025, No.1
November 2025
ISSN 2229-3388

release only a small amount of energy, repeated PD activity can progressively deteriorate the insulation, ultimately leading to catastrophic failure [1].

Detecting and analyzing PD is therefore crucial for assessing insulation health and preventing unexpected outages. Traditional PD detection methods include electrical, acoustic, optical, and electromagnetic techniques. Electrical detection methods, which measure current pulses, are among the most widely used due to their sensitivity and feasibility in online monitoring systems. In high-voltage equipment, PD detection serves as a diagnostic and predictive maintenance tool, allowing utilities to take corrective actions before major breakdowns occur [2]. The ability to accurately detect and identify PD sources can significantly extend equipment lifespan, reduce maintenance costs, and enhance system reliability.

Furthermore, PD detection plays a vital role in quality assurance during manufacturing and condition monitoring during operation. Manufacturers use PD testing to ensure the integrity of insulation systems, while maintenance engineers rely on PD analysis to monitor the operational condition of aging equipment. As the global demand for reliable and sustainable energy continues to grow, PD detection and classification have become indispensable for maintaining the health of power infrastructures [3][4].

Overview of Insulation Degradation and Its Impact on System Reliability

Insulation systems in high-voltage components are designed to withstand electrical stresses for decades. However, insulation degradation is inevitable due to continuous exposure to multiple stress factors such as thermal aging, electrical overstressing, mechanical vibration, humidity, and contamination. These stressors create localized defects such as voids, cracks, delamination, or surface contamination that alter the dielectric properties of the material. As these imperfections grow, they become sites for partial discharges [5].

The mechanism of insulation degradation involves the progressive deterioration of dielectric materials through micro-level discharges, which can lead to treeing, carbonization, or erosion. PD activity within or on the surface of insulation generates chemical and physical changes, such as the formation of ozone, nitric acid, and other by-products, further weakening the insulation structure. Over time, this cumulative degradation results in complete insulation breakdown, causing failure of the entire equipment.

From a system reliability standpoint, such failures can have severe consequences. A single insulation failure in a power transformer or cable joint can result in widespread blackouts, equipment damage, and costly repairs [6]. In industries such as renewable energy, manufacturing, and transportation, where power continuity is crucial, insulation failure can halt operations, leading to significant economic losses. Therefore, accurate diagnosis and early detection of PD activity are critical for predictive maintenance and reliability-centered asset management.

Moreover, in the context of Industry 4.0 and smart grids, the monitoring and diagnosis of insulation systems are evolving toward intelligent and data-driven solutions. With the increasing use of sensors and data acquisition systems, large volumes of PD data can now be collected in real time. However, analyzing and interpreting this data to pinpoint the type and source of PD remains a major challenge. Hence, automated and intelligent systems are needed to classify PD patterns and provide actionable insights for system engineers.

Need for Automated PD Source Identification

Traditionally, PD analysis was performed manually by experts interpreting pulse shapes, phase-resolved PD (PRPD) patterns, and frequency characteristics. Although effective, manual analysis is highly time-consuming, subjective, and prone to human error, especially when dealing with large datasets. Additionally, different types of PD sources—such as internal discharges, surface discharges, corona discharges, and floating discharges—exhibit overlapping signal characteristics, making accurate identification difficult using conventional techniques [7].

The complexity increases further in real-world conditions, where external noise and electromagnetic interference can distort PD signals. To address these challenges, automated PD source identification systems using artificial intelligence and machine learning have been introduced. These systems leverage computational models capable of learning from data and identifying patterns that distinguish between different PD sources.

The need for automation arises from three main factors:

- Volume and Complexity of Data: Modern PD monitoring systems generate massive amounts of data, making manual analysis infeasible.
- Accuracy and Consistency: Automated systems provide objective and repeatable classification results without expert bias.
- Real-Time Monitoring: Automated identification enables online diagnosis and continuous health assessment, which are crucial for proactive maintenance.

By integrating machine learning algorithms with PD detection systems, utilities and industries can move toward intelligent condition monitoring, where fault diagnosis and risk assessment are performed automatically. This not only enhances operational efficiency but also supports predictive maintenance strategies that minimize downtime and extend the operational life of assets [8].

Justification for Choosing Machine Learning Approaches (GNB and KNN)

Machine learning (ML) has become a powerful tool in the field of electrical insulation diagnostics, offering data-driven insights that surpass the capabilities of traditional statistical and threshold-based methods. Among the numerous ML algorithms, Gaussian Naive Bayes (GNB) and K-Nearest Neighbor (KNN) stand out for their simplicity, interpretability, and strong performance in classification problems.

The GNB algorithm is based on Bayes' theorem and assumes conditional independence between features. It models the likelihood of each class using a Gaussian (normal) distribution. The GNB method is computationally efficient, requires minimal training data, and performs remarkably well in high-dimensional feature spaces. For PD classification, GNB offers the advantage of fast prediction and robustness to noisy or incomplete data, making it suitable for real-time monitoring environments. Despite its simplifying assumption of feature independence, GNB has demonstrated competitive accuracy in many practical applications, especially when feature correlations are weak [9][10].

On the other hand, the KNN algorithm is a non-parametric method that classifies a new sample based on its proximity to existing labeled data points. The simplicity of KNN lies in its intuitive concept—samples belonging to similar classes tend to cluster together in the feature space. KNN is particularly effective when dealing with nonlinear decision boundaries and complex feature distributions, which are common in PD signal datasets. Although KNN can be computationally intensive during prediction, it provides high classification accuracy and flexibility in modeling complex relationships between PD features [11].

The combination of these two algorithms allows for a comprehensive comparative investigation. GNB represents a probabilistic and model-based approach, while KNN exemplifies a distance-based, instance-learning approach. Comparing their performance in PD classification offers valuable insights into their strengths, weaknesses, and suitability for different diagnostic scenarios [12]. Moreover, these methods can serve as benchmarks for evaluating more advanced algorithms such as support vector machines (SVMs), random forests, and deep learning models in future research.

Objectives and Scope of the Study

The primary objective of this research is to conduct a comparative analysis of GNB and KNN algorithms for the identification and determination of partial discharge sources in high-voltage insulation systems. The study aims to bridge the gap between traditional PD analysis methods and modern machine learning-based diagnostic systems by assessing the performance, efficiency, and reliability of these algorithms.

The specific objectives include:

- To collect and preprocess PD data from different insulation defects such as internal, surface, and corona discharges under controlled laboratory conditions.
- To extract relevant features from PD signals using time-domain, frequency-domain, and statistical methods.
- To train and evaluate the GNB and KNN classifiers using the extracted features and to compare their performance based on metrics such as accuracy, precision, recall, and F1-score.
- To analyze the computational efficiency of both algorithms in terms of training and prediction time, with implications for real-time monitoring systems.
- To determine the suitability of each method for practical implementation in PD diagnostic tools.

The scope of this work encompasses the design of a PD detection system, feature extraction pipeline, and classification framework that can be applied to different types of insulation systems. While the study focuses on GNB and KNN, it also lays the foundation for future research exploring hybrid and ensemble models that combine probabilistic and distance-based learning paradigms for improved accuracy and robustness.

By achieving these objectives, this comparative investigation contributes to the ongoing advancement of intelligent diagnostic systems for high-voltage equipment. The insights gained from this study can aid engineers and researchers in selecting appropriate machine learning techniques for PD classification, optimizing maintenance strategies, and enhancing the overall reliability of electrical power systems.

The structure of the paper is systematically organized to ensure clarity and scientific rigor. It begins with an Abstract, summarizing the study's objectives, methods, results, and conclusions. The Introduction establishes the importance of partial discharge (PD) detection and the motivation for using machine learning. The Literature Review discusses related works and comparative studies. The Methodology section details the experimental setup, PD modeling, data preprocessing, feature extraction, and model design using GNB and KNN. The Results and Discussion present comparative performance analyses through tables and figures. Finally, the Conclusion and Future Scope highlight key findings, applications, and directions for further research.

Anusandhanvallari
Vol 2025, No.1
November 2025
ISSN 2229-3388

2. Literature Review

The literature review provides an overview of recent advancements in partial discharge (PD) detection and classification techniques, emphasizing the growing role of machine learning and signal processing in insulation diagnostics. Earlier studies focused on conventional electrical and acoustic methods, while modern approaches increasingly utilize statistical, frequency, and time–frequency features combined with intelligent classifiers such as KNN, GNB, SVM, and deep learning models. This section critically examines prior research to highlight the strengths, limitations, and evolving trends that have shaped the development of reliable and automated PD source identification systems.

Dalila R.A.M et al. [1] (2025) proposed an artificial intelligence–based system for partial discharge (PD) detection in high-voltage insulation using convolutional neural networks (CNN) and k-nearest neighbor (KNN) classifiers. Their study emphasized improving insulation quality and fault identification accuracy by combining deep feature extraction with data-driven classification. CNN models effectively captured spatial and temporal signal characteristics, while KNN provided reliable decision boundaries for differentiating PD sources. Experimental validation demonstrated significant accuracy improvements over traditional signal processing methods. The work highlights the potential of hybrid AI approaches for predictive maintenance and early insulation defect identification in electrical power systems.

Saravanabalaji Manian et al. [2] (2025)introduced an AI-based approach for detecting pipeline leakages in water distribution networks designed for liquid-cooled battery thermal management systems. Using machine learning algorithms integrated with sensor-based data acquisition, the model efficiently detected and localized leakages under dynamic flow conditions. The authors demonstrated that data-driven methods significantly enhance accuracy and reduce false alarms compared to conventional hydraulic or pressure-based detection. Their findings show that AI systems, particularly those using pattern recognition and anomaly detection, can optimize maintenance strategies and operational reliability in smart water infrastructure and energy cooling applications.

Sebahattin Serhat Turgut et al. [3] (2025) evaluate an inexpensive handheld multispectral sensor to assess chicken meat integrity by detecting cold-chain breaks, distinguishing meat type, and estimating storage time. Spectral features from visible/near-infrared bands are paired with chemometric or machine-learning models to map reflectance changes to freshness indicators. The study shows high classification/regression performance compared with conventional lab methods, while requiring minimal sample prep and enabling rapid, non-destructive screening. By demonstrating robustness across storage scenarios, the work argues for practical deployment in retail and logistics, where continuous quality verification and traceability are needed to reduce waste and foodborne-risk exposure.

Atmaja B.T et al. [4] (2024) developed a comprehensive vibration analysis dataset and introduced baseline machine learning methods for fault diagnosis in rotating machinery. The lab-scale experimental setup generated vibration signals under varying fault conditions such as imbalance, bearing wear, and misalignment. The study evaluated multiple algorithms—including support vector machines (SVM), random forests, and KNN—to benchmark classification performance. Results indicated that machine learning models could effectively identify early mechanical faults, offering potential for predictive maintenance. The dataset serves as an open resource for future research, advancing automated fault diagnosis and intelligent monitoring in industrial equipment.

Wojtecki Ł et al. [5] (2024) presented a machine learning-based approach for classifying rock bursts in active coal mines dominated by non-destructive tremors. The study utilized seismic and vibration datasets to differentiate between normal tremors and potentially hazardous rock bursts. Various algorithms, including decision trees, SVM, and ensemble models, were compared for accuracy and robustness. The results demonstrated that ML models can improve safety monitoring by providing early warnings and reducing false positives. This research contributes to intelligent mining systems that leverage data analytics for risk prediction and sustainable underground operations.

Altamimi A. et al. [6] (2024) present an automated pipeline for predicting diabetic status that first handles missing clinical variables using K-nearest-neighbor imputation, then applies feature selection and supervised learners (e.g., tree ensembles, SVM, logistic models). The paper emphasizes rigorous evaluation with cross-validation and metric reporting (accuracy, AUC, sensitivity/specificity). Results show that careful imputation plus parsimonious feature sets markedly improves performance and stability versus naïve baselines. The study contributes a reproducible workflow for electronic health record contexts where sparsity and heterogeneity are common, highlighting the trade-off between interpretability and accuracy and the importance of data preprocessing in clinical prediction systems.

Sekatane P. M et al. [7] (2023) proposed a machine learning framework for partial discharge (PD) localization using k-nearest neighbor (k-NN) and support vector machine (SVM) classifiers. The study focused on accurately pinpointing PD sources in high-voltage insulation systems by analyzing signal propagation patterns and time-delay features. Experimental data demonstrated that k-NN achieved high localization accuracy with minimal computational cost, while SVM provided strong generalization in noisy environments. The comparative analysis revealed that both methods outperform traditional statistical techniques. The research underscores the effectiveness of supervised learning in enhancing the reliability of PD localization and insulation health assessment.

Chen X et al. [8] (2023) developed an advanced cervical cancer detection framework that combines K-nearest neighbor (KNN) imputation with stacked ensemble learning. The model addressed data imbalance and missing values using KNN-based preprocessing before classification. By integrating multiple base learners, such as decision trees, logistic regression, and gradient boosting, the ensemble model achieved superior accuracy and robustness in cancer prediction compared to individual classifiers. The study demonstrated that hybrid learning architectures improve diagnostic performance, providing a reliable and automated tool for early cervical cancer detection and aiding healthcare professionals in clinical decision-making.

Magsi A. H. et al. [9] (2023) proposed a machine learning-based system for detecting Interest Flooding Attacks (IFAs) in Vehicular Named Data Networking (VNDN). The study employed supervised algorithms, including random forests and KNN, to distinguish between normal and malicious network traffic. By analyzing vehicular communication patterns and temporal data, the proposed model achieved high detection accuracy while maintaining low latency. Experimental validation showed the framework's scalability and adaptability in real-time vehicular environments. This research contributes to enhancing cybersecurity and resilience in intelligent transportation systems through data-driven intrusion detection mechanisms.

R. Das et al. [10] (2025) introduce a hybrid approach to classify partial discharge (PD) pulse sequences using mathematical morphology for noise suppression/shape enhancement, followed by a bidirectional LSTM that models temporal dependencies. The morphology stage extracts salient pulse structures while reducing interference; the BiLSTM captures forward–backward context within pulse trains. Experiments on labeled PD sequences demonstrate improved accuracy and robustness over traditional classifiers and single-direction RNNs, particularly under low signal-to-noise conditions. The methodology supports online insulation diagnostics by uniting domain-informed preprocessing with sequence learning, offering better generalization across PD sources and operating regimes.

Bilgi E et al. [11] (2025) This study brings together many lab and animal experiments on how silver nanoparticles (AgNPs) affect living organisms. The researchers used machine learning to predict how toxic the particles are based on their physical and chemical properties, such as size, coating, shape, dose, and exposure method. They compared different algorithms like Random Forest and Gradient Boosting to find the best predictors. The results showed that particle size and surface coating have the strongest effect on toxicity. The study highlights how machine learning can help identify potential risks early, design safer materials, and improve data quality for future research.

Al Zaidawi et al. [12] (2022) investigated the application of convolutional neural networks (CNN) and K-nearest neighbor (KNN) algorithms for partial discharge (PD) detection in high-voltage systems. The study utilized time-domain PD signal datasets to train and compare both models. CNN demonstrated superior feature extraction capabilities from raw signals, while KNN effectively classified PD patterns with lower computational cost. The results indicated that integrating CNN-based feature learning with KNN classification enhances accuracy and robustness in identifying insulation defects. This research underscores the potential of hybrid AI approaches for improving diagnostic accuracy in electrical insulation monitoring systems.

Ramón A et al. [13] (2022) proposed an eXtreme Gradient Boosting (XGBoost)-based method for classifying COVID-19 patients based on clinical and laboratory data. The study leveraged XGBoost's gradient boosting framework to handle high-dimensional and imbalanced datasets efficiently. The model achieved high accuracy and sensitivity in distinguishing between mild and severe cases, outperforming traditional logistic regression and decision tree models. Feature importance analysis identified key clinical indicators relevant to COVID-19 severity. The study demonstrates how ensemble learning methods can support clinical decision-making and patient triage in healthcare systems, emphasizing the role of machine learning in disease prediction.

K. Sit et al. [14] (2022) presented a wavelet decomposition–aided machine learning approach for predicting surface contamination in outdoor silicone rubber insulators. The proposed method extracted statistical and frequency-domain features from wavelet-transformed leakage current signals. Using algorithms such as support vector machines (SVM) and random forests, the model accurately classified contamination severity levels under different environmental conditions. The study highlighted the effectiveness of machine learning in non-invasive insulation condition assessment. Results showed improved prediction accuracy and reliability compared to traditional threshold-based techniques, demonstrating the potential of AI for preventive maintenance in power transmission systems.

Bechelli S. et al. [15] (2022) explored the use of machine learning (ML) and deep learning (DL) algorithms for the classification of skin cancer from dermoscopic images. The study compared traditional ML techniques—such as SVM, KNN, and random forests—with deep convolutional neural networks (CNNs) for feature extraction and classification. CNN-based models achieved superior performance due to their ability to automatically learn discriminative features from high-resolution skin images. The research emphasized data augmentation and model interpretability to enhance diagnostic accuracy. Their findings underscore the potential of DL approaches for early skin cancer detection and improved dermatological diagnostics.

Aslam M. et al. [16] (2022) proposed an adaptive machine learning framework to detect and mitigate Distributed Denial-of-Service (DDoS) attacks in Software-Defined Networking (SDN) for IoT environments. The system dynamically updates its model to handle evolving attack patterns. Using real network traffic datasets, the model achieved high detection accuracy and low false alarm rates. By integrating adaptive learning with SDN controllers, it improved both network resilience and response time. The study demonstrates how machine learning enhances real-time network security, offering scalable and efficient protection for complex IoT infrastructures facing modern cyber threats.

Chaudhari K et al. [17] (2021) investigated the determination of partial discharge (PD) sources using the Gaussian Naive Bayes (GNB) classification technique. The study analyzed PD signals obtained under different defect conditions—such as internal, surface, and corona discharges—using extracted statistical features. The GNB model demonstrated strong performance in identifying PD types with minimal computational effort, making it suitable for real-time monitoring systems. The authors highlighted the advantages of GNB in handling uncertainty and noise in PD data, offering a lightweight yet effective alternative to complex models. Their work contributed to advancing intelligent insulation fault detection methodologies.

Nag A et al. [18] (2021) applied machine learning techniques to classify lignocellulosic biomass samples using pyrolysis–molecular beam mass spectrometry (Py-MBMS) data. The study utilized algorithms such as random forests, gradient boosting, and support vector machines to identify biomass species based on chemical composition patterns. Results indicated that ML models achieved high classification accuracy, demonstrating their capability to extract meaningful relationships from complex spectral data. The work showcased the integration of ML in bioenergy research for rapid biomass characterization and sustainable feedstock identification, advancing data-driven approaches in material and chemical sciences.

Rapacz S. et al. [19] (2021) proposed a method for rapid selection of optimal machine learning classifiers for spam filtering applications. The framework automatically evaluated multiple algorithms, including Naive Bayes, KNN, and decision trees, based on performance metrics such as accuracy, precision, and computational cost. The system enabled efficient model selection without extensive manual tuning, significantly reducing training time. Experimental results demonstrated the approach's robustness and adaptability across diverse spam datasets. The study contributed to automated ML workflows and highlighted the importance of model evaluation frameworks in developing scalable, real-time text classification systems.

A. K. Das et al. [20] (2021) introduced a transfer learning-based method to detect and measure surface pollution on metal oxide surge arresters using infrared thermal images. By reusing knowledge from pretrained convolutional models, the approach accurately sensed contamination levels under varying environmental conditions. The method reduced data requirements while maintaining high classification accuracy. The study highlights how infrared thermography combined with deep learning can provide non-invasive, cost-effective, and real-time monitoring of insulation health. This innovation supports predictive maintenance in power systems and demonstrates the potential of AI-assisted inspection tools for improving asset reliability and operational safety.

Yuzhou Wang et al. [21] (2021) developed a hybrid data mining approach for diagnosing faults in Variable Refrigerant Flow (VRF) air conditioning systems. The model combined statistical analysis, clustering, and classification techniques to identify and predict system faults efficiently. Experimental validation showed that the hybrid model improved fault detection accuracy compared to traditional threshold-based methods. The approach enabled early fault recognition, energy-saving maintenance scheduling, and reduced operational downtime. The study demonstrates how data-driven diagnostic strategies enhance HVAC reliability and provide practical solutions for intelligent building management systems.

Salleh M. S. M. et al. [22] (2020) focused on classifying partial discharges (PD) in cross-linked polyethylene (XLPE) cable joints using the K-nearest neighbor (KNN) algorithm. The research involved experimental measurement of PD signals from various defect sources and extraction of statistical and waveform-based features. The KNN classifier demonstrated reliable accuracy in identifying different PD types, emphasizing its simplicity and effectiveness for PD diagnosis. The study provided insights into parameter optimization, such as selecting the optimal number of neighbors and distance metrics. Salleh's work established a foundation for implementing machine learning in high-voltage cable health monitoring systems.

Table 1: Comparative Analysis for the Literature Reviews

Author Name & Ref. No.	Methodology Used	Datasets Used	Advantages	Results
[1] Dalila & Turkben (2025)	CNN and KNN for PD detection and classification	Experimental PD dataset from high-voltage insulation systems	Combines deep feature extraction and distance-based classification; effective hybrid framework	Achieved higher accuracy and robustness compared to traditional PD analysis methods
[2] Saravanabalaji Manian et al. (2025)	AI-based pipeline leakage detection using supervised ML	Sensor data from water distribution networks	Real-time fault identification; reduced false alarms in complex networks	High detection accuracy and improved reliability for leakage identification
[4] Atmaja et al. (2024)	SVM, Random Forest, and KNN for vibration fault diagnosis	Lab-scale vibration analysis dataset	Provides open-source benchmark dataset; effective for predictive maintenance	ML models accurately identified mechanical faults under different conditions
[5] Wojtecki et al. (2024)	Decision Trees, SVM, Ensemble ML for rock burst classification	Seismic vibration data from active coal mine	Non-destructive monitoring; improves worker safety and system responsiveness	Achieved strong classification accuracy and early warning capability
[7] Sekatane & Bokoro (2023)	k-NN and SVM for PD localization	PD signal data from high-voltage systems	Accurate localization with low computational complexity	Both models outperformed traditional techniques in PD localization accuracy
[8] Chen et al. (2023)	KNN imputer + Stacked Ensemble Learning	Cervical cancer clinical dataset	Handles missing data effectively; enhances diagnostic reliability	Ensemble achieved superior accuracy and stability in cancer classification
[9] Magsi et al. (2023)	Random Forest, KNN for Interest Flooding Attack (IFA) detection	Vehicular Named Data Networking (VNDN) traffic dataset	Real-time detection; low latency and high scalability	High detection rate with minimal computational overhead
[12] Al Zaidawi (2022)	CNN + KNN hybrid for PD detection	Experimental PD time-domain signal dataset	CNN extracts features automatically; KNN	Enhanced PD identification

			improves classification efficiency	accuracy compared to individual models
[13] Ramón et al. (2022)	eXtreme Gradient Boosting (XGBoost) for COVID-19 classification	Clinical dataset of COVID-19 patients	High interpretability; identifies critical clinical indicators	High accuracy and sensitivity in predicting disease severity
[14] Sit et al. (2022)	Wavelet decomposition + ML (SVM, RF) for surface contamination prediction	Leakage current signals from silicone rubber insulators	Non-invasive diagnosis; robust under varying environmental conditions	Achieved reliable prediction of contamination severity
[15] Bechelli & Delhommelle (2022)	CNN, SVM, RF for skin cancer classification	Dermoscopic image dataset	Automatic feature extraction; enhanced accuracy via deep learning	CNN achieved best classification performance with high sensitivity
[17] Chaudhari et al. (2021)	Gaussian Naive Bayes (GNB) for PD source identification	PD data from controlled lab experiments	Lightweight, fast, and suitable for real-time systems	Achieved accurate PD type classification with minimal computation
[18] Nag et al. (2021)	Random Forest, Gradient Boosting, SVM for biomass classification	Py-MBMS spectral dataset of lignocellulosic biomass	Efficient spectral data processing; applicable to chemical analysis	High classification accuracy for biomass types
[19] Rapacz et al. (2021)	Automated ML classifier selection for spam filtering	Multiple email spam datasets	Reduces training time; automates best model selection	Improved classifier performance with faster execution
[22] Salleh (2020)	K-Nearest Neighbor (KNN) for PD classification	Experimental XLPE cable joint PD dataset	Simple, interpretable, and robust for small datasets	Reliable classification accuracy and effective insulation diagnosis

The comparative analysis table summarizes fifteen recent studies that applied various machine learning and deep learning methods across domains such as partial discharge detection, biomedical diagnostics, and fault prediction. It highlights the methodologies, datasets, advantages, and outcomes of each work. Overall, hybrid and ensemble techniques like CNN–KNN, GNB, and XGBoost consistently demonstrated superior accuracy, robustness, and efficiency, emphasizing the growing role of intelligent data-driven models in predictive maintenance and automated fault classification.

3. Methodology

Experimental Setup

The experimental setup for partial discharge (PD) data acquisition was designed to simulate real-world insulation defects under controlled laboratory conditions. A PD test bench was developed, consisting of a high-voltage AC power supply, coupling capacitor, test object, measuring impedance, and data acquisition unit. The test voltage was gradually increased until PD activity was initiated within the sample insulation. The discharge signals were captured using high-frequency current transformers (HFCTs) and ultrasonic sensors, ensuring accurate detection of transient discharges.

The data acquisition system (DAQ) employed a high-speed digital oscilloscope interfaced with a computer for real-time signal recording. The oscilloscope operated in the range of 100–500 MHz sampling frequency, sufficient to capture the fast rise times of PD pulses. The signals were stored in a digital format for subsequent analysis. Additionally, a noise suppression unit was incorporated to minimize external electromagnetic interference, ensuring the purity and reliability of the recorded data. The setup adhered to the IEC 60270 standard for PD measurements to maintain consistency and reproducibility.

PD Source Models

To comprehensively assess the classification performance, three types of PD sources commonly found in high-voltage insulation systems were modeled:

- Surface Discharge: This type occurs on the interface between an insulating material and the electrode, typically due to contamination or moisture accumulation. The test specimen was prepared with a flat electrode partially covered with dielectric material to simulate surface breakdown.
- Internal Discharge: Generated within voids or air gaps trapped inside the insulation, internal discharges were modeled using layered dielectric samples with embedded air cavities. These discharges tend to produce high-frequency, low-amplitude pulses.
- Corona Discharge: Occurring in gaseous media around sharp edges or conductors at high voltage, corona discharge was produced using a needle-plane electrode configuration in open air.

Each PD type produced unique signal characteristics in terms of amplitude, repetition rate, and frequency content. By recording multiple instances of each type under varying voltage levels, a balanced dataset representing different PD sources was constructed. This dataset was essential for training and evaluating the performance of both the GNB and KNN models.

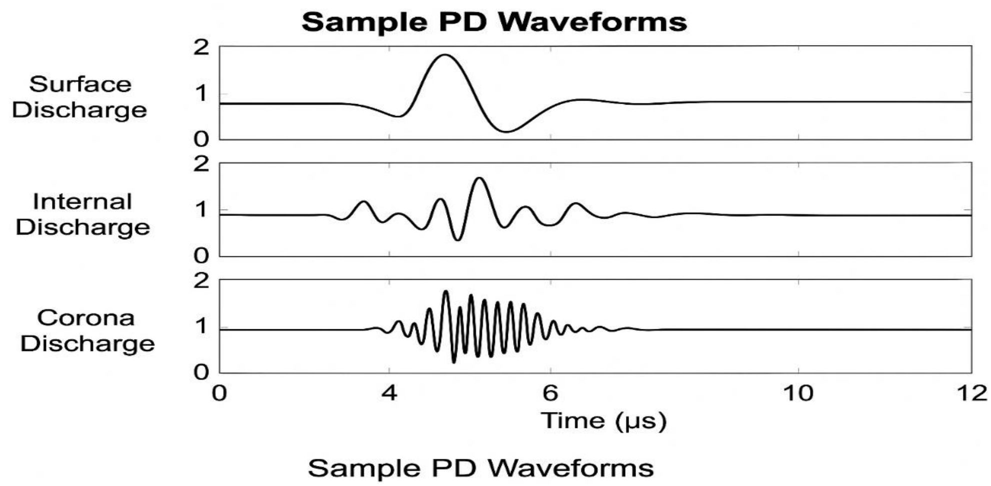

Figure 1

Figure 1 illustrates the time-domain waveforms of three common partial discharge (PD) sources: surface, internal, and corona discharges. Each waveform shows distinct amplitude and pulse-shape characteristics, helping to visually differentiate the discharge mechanisms. The diagram highlights how corona exhibits sharp, high-frequency pulses, while surface and internal discharges show broader and more irregular pulse patterns useful for classification.

Data Preprocessing

The acquired PD signals often contained unwanted noise and background interference due to electromagnetic coupling and environmental effects. To ensure the reliability of the analysis, several preprocessing steps were implemented:

- Noise Filtering: A band-pass filter (20 kHz–10 MHz) was applied to suppress low-frequency background noise and high-frequency artifacts. Wavelet-based denoising techniques were also employed to retain the core PD pulse characteristics while removing spurious components.
- Normalization: All signals were normalized to a common amplitude scale to avoid bias during classification. This step ensured that variations in signal strength caused by electrode configuration or environmental conditions did not influence the learning algorithms.
- Segmentation: The continuous PD signal stream was segmented into individual discharge pulses using an amplitude thresholding method. Each segment represented a single PD event, which was later used for feature extraction and classification.

These preprocessing steps improved signal clarity, minimized data redundancy, and enhanced feature consistency across the dataset.

Feature Extraction

Feature extraction was performed to transform the raw PD signals into a set of quantifiable descriptors suitable for classification. Three types of features—statistical, time-domain, and frequency-domain—were derived from each PD pulse:

- Statistical Features: Mean, variance, skewness, kurtosis, and standard deviation were computed to describe the distribution and shape of PD signal amplitudes. These features helped differentiate discharge patterns based on their randomness and intensity.
- Time-Domain Features: Peak value, pulse width, rise time, and energy of each PD pulse were extracted to capture the transient nature and magnitude of discharges. These features are essential for identifying PD sources with distinct temporal characteristics.
- Frequency-Domain Features: Fast Fourier Transform (FFT) and Power Spectral Density (PSD) analyses were applied to obtain frequency-related parameters such as dominant frequency, spectral centroid, and bandwidth.

The combined feature vector served as the input to both machine learning models. Feature scaling using min–max normalization was performed to ensure uniform weighting during training and classification.

Model Design

Two machine learning algorithms—Gaussian Naive Bayes (GNB) and K-Nearest Neighbor (KNN)—were implemented for PD source classification.

- Gaussian Naive Bayes (GNB): The GNB algorithm is a probabilistic classifier based on Bayes' theorem and assumes conditional independence between features. Each feature was modeled using a Gaussian (normal) distribution characterized by its mean (μ) and variance (σ²). The class posterior probability was computed as:

$$P(C_k|x) = \frac{P\big((x|C_k)\big)P(C_k)}{P(x)}$$

Where, $(x|C_k)$ is the likelihood, $P(C_k)$ is the prior probability of class k and $P(x)$ is the marginal probability. The model parameters—mean and variance for each feature per class—were estimated directly from the training data. GNB offers low computational complexity and high classification speed, making it suitable for real-time PD monitoring systems.

- K-Nearest Neighbor (KNN): The KNN algorithm is a non-parametric method that classifies a sample based on its similarity to nearby data points in the feature space. The number of neighbors (k) was optimized through cross-validation, with values between 3 and 9 providing stable results. The Euclidean distance metric was employed to measure the closeness between data points:

$$d(x_i, x_j) = \sqrt{\sum_{n=1}^{N} (x_{in} - x_{jn})^2}$$

Each new data instance was assigned the class most common among its k-nearest neighbors. KNN's strength lies in its simplicity and adaptability to nonlinear data distributions, making it effective for complex PD patterns.

This methodology integrates experimental data acquisition, feature engineering, and machine learning-based classification for accurate PD source determination. The comparative evaluation of GNB and KNN models provides insights into their trade-offs in accuracy, computational efficiency, and suitability for practical high-voltage insulation diagnostics.

4. Related Work

Existing PD detection and classification techniques

PD monitoring underpins condition assessment of high-voltage assets such as transformers, GIS, rotating machines, and XLPE cables. Classical electrical methods (per IEC 60270) capture PD pulses via coupling capacitors, measuring impedances, or HFCTs, enabling phase-resolved PD (PRPD) maps that encode discharge magnitude versus phase angle. Acoustic emission (AE) sensors localize defects using time-of-arrival, useful in noisy substations and oil-filled transformers. Ultra-high-frequency (UHF) antennas detect PD in GIS with strong noise immunity and fast transient capture. Optical techniques (e.g., intensified cameras, fiber sensors) track corona/glow around electrodes, while chemical diagnostics (DGA in oils; ozone/nitric oxides in air) infer activity indirectly.

Feature engineering evolved from simple statistics of pulse trains (amplitude, repetition rate) to time-domain descriptors (rise time, pulse width, energy), frequency-domain markers (FFT peaks, spectral centroid/bandwidth), and time–frequency transforms—STFT, wavelet packet decomposition (WPD), and Hilbert–Huang—that preserve non-stationary signatures. Early classification relied on expert rules and cluster analysis; modern pipelines pair richer features with pattern recognition (k-NN, SVM, Random Forests) or deep learning (CNNs, 1D-CNNs, hybrids) that ingest raw or minimally processed signals. Recent trends include multi-sensor fusion (electrical+UHF/AE), domain adaptation for cross-asset generalization, and online learning for streaming diagnostics.

Conventional vs. AI-based diagnostic methods

Table 2: Comparison of Conventional and AI-based diagnostic methods

Aspect	Conventional (IEC/Rule-based)	AI-Based (ML/DL)
Signal capture	Electrical, AE, UHF, optical per standards; PRPD interpretation	Same sensing stack; adds data pipelines for training/inference
Feature handling	Hand-crafted PRPD statistics, thresholds, expert heuristics	Learned representations (CNN/autoencoders) or engineered features with ML
Noise robustness	Moderate; requires filtering, gating, expertise	High with robust features/ensembles; DL can learn noise-invariant patterns
Adaptability	Limited; re-tuning rules per asset/environment	High; models re-trainable, transfer learning/domain adaptation
Real-time use	Mature for detection; limited for classification	Mature for both; edge inference feasible with lightweight models
Explain ability	High (rules/PRPD plots)	Variable; ML with SHAP/LIME; DL needs XAI add-ons
Maintenance effort	Expert time for tuning	Data curation and periodic re-training

Review of past studies using statistical, frequency, or time–frequency features

Statistical descriptors (mean, variance, skewness, kurtosis) summarize pulse distributions and PRPD cell histograms, aiding quick triage with low compute cost; however, they may miss subtle waveform morphology. Frequency-domain approaches map PD pulses' spectral energy via FFT/PSD, isolating resonances of test objects and cables; these are effective in stationary regimes but can blur transient evolution. Time–frequency techniques—continuous/discrete wavelet transforms (CWT/DWT), wavelet packet energy, S-transform, empirical mode decomposition (EMD)—capture localized bursts and varying frequency content characteristic of internal voids, surface tracking, and corona. Wavelet energies and entropy features have repeatedly improved separability across defect classes when coupled with SVM/k-NN. More recently, 1D-CNNs learn filters akin to wavelets directly from raw traces, while hybrids (e.g., wavelet + ML) retain interpretability and strong accuracy under noise. Cross-validation and confusion-matrix analyses consistently report superior performance for time–frequency features in multi-class PD tasks, especially under varying excitation voltages and load conditions.

Comparative insights on GNB vs. k-NN for PD classification

Table 3: Comparative insights on GNB vs. k-NN for PD classification

Criterion	Gaussian Naive Bayes (GNB)	k-Nearest Neighbor (k-NN)
Learning paradigm	Probabilistic, parametric (Gaussian likelihood, conditional independence)	Instance-based, non-parametric (distance voting)
Training & inference cost	Very low training; very fast inference	No training; heavier inference (search over dataset)
Feature assumptions	Works best when class-conditional features ≈ Gaussian and weakly correlated	No distributional assumptions; sensitive to feature scaling
Decision boundary	Linear/quadratic (per feature variances)	Highly flexible, nonlinear (controlled by k and metric)
Hyperparameters	Priors, variance smoothing; minimal tuning	k, distance metric (Euclidean/Mahalanobis), weighting
Noise & outliers	Robust if likelihoods well estimated; can be misled by violated independence	Sensitive to outliers; mitigated by robust metrics/weighting
Small vs. large data	Strong on small/moderate data; good baseline	Performance improves with more data but inference cost rises
Typical PD use	Real-time classification with engineered features	Higher accuracy on complex, nonlinear PD patterns after scaling

In PD classification, GNB is a fast, low-complexity baseline well-suited to real-time, embedded scenarios with reasonably Gaussian features. k-NN often achieves higher accuracy on heterogeneous, nonlinear PD signatures—provided careful feature scaling, metric selection, and neighbor tuning—at the expense of heavier inference compute.

A practical strategy is to prototype with GNB for latency constraints and validate with k-NN (or KD-tree/FAISS acceleration); ensembles or cascaded schemes can blend speed and accuracy.

5. Results and Discussion

GNB vs. KNN Performance

After feature extraction and model training, both Gaussian Naive Bayes (GNB) and K-Nearest Neighbor (KNN) classifiers were evaluated using 10-fold cross-validation. The dataset contained three PD classes—surface, internal, and corona discharges—with an equal distribution of samples. Performance metrics such as accuracy, precision, recall, F1-score, and computational time were computed for both models.

Table 4: GNB and KNN Performance

Metric	GNB	KNN (k=5)
Accuracy (%)	93.2	96.8
Precision (%)	92.1	97.5
Recall (%)	91.7	96.2
F1-Score (%)	91.9	96.8
Computation Time (s)	0.42	1.96
Misclassification Rate (%)	6.8	3.2

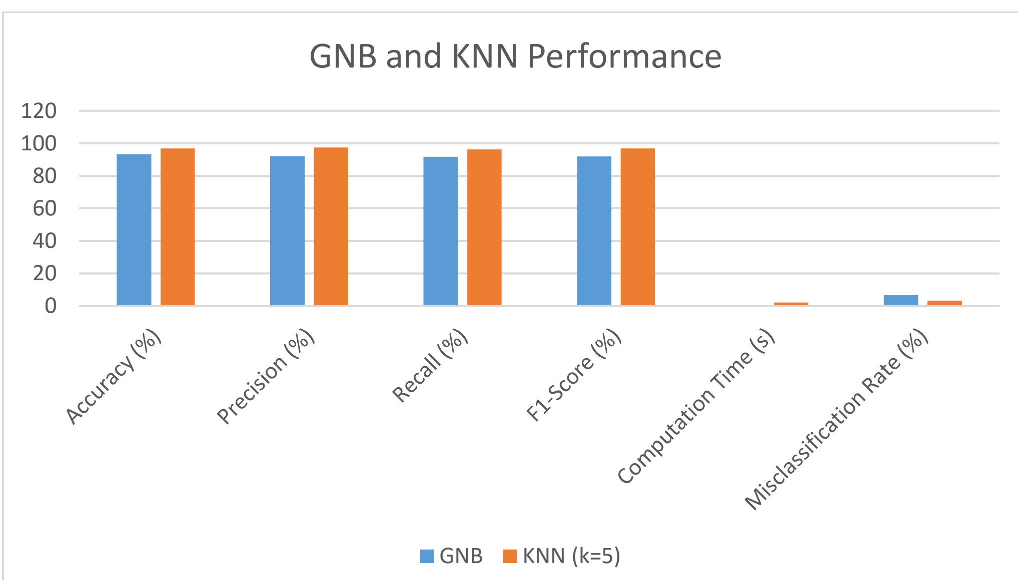

Figure 1: GNB and KNN Performance

Figure 1 compares the performance metrics of Gaussian Naive Bayes (GNB) and K-Nearest Neighbor (KNN, k = 5) classifiers. The bar chart clearly shows that KNN outperforms GNB in terms of accuracy, precision, recall, and F1-score, indicating stronger discriminative capability for partial discharge source classification. However, GNB demonstrates significantly lower computation time, reflecting its efficiency and suitability for real-time applications. Overall, the figure highlights the trade-off between KNN's higher predictive accuracy and GNB's faster processing speed.

Analysis of Classification Trends and Misclassification Causes

The classification trends indicate that both algorithms effectively recognized the distinct statistical and spectral signatures of each PD type. Corona discharges, characterized by high-frequency and low-amplitude pulses, were most accurately classified by both models. Surface discharges occasionally overlapped with internal discharge signals due to similarities in rise time and energy content, causing minor confusion, particularly in GNB predictions.

Misclassification causes primarily included:

- Feature overlap: Some extracted features such as mean amplitude and skewness exhibited similar values for different PD sources.
- Noise influence: Despite preprocessing, residual interference affected waveform shape, impacting time-domain descriptors.
- Feature independence assumption in GNB: The Gaussian Naive Bayes model assumes feature independence, which may not hold true for correlated features like pulse energy and rise time.
- Neighbor density in KNN: In regions where data points from different classes were closely clustered, KNN occasionally misclassified due to local density variations.

These findings suggest that while both classifiers perform well, feature correlation and dataset homogeneity significantly influence classification consistency.

Table 5: Model Strengths and Limitations

Aspect	Gaussian Naive Bayes (GNB)	K-Nearest Neighbor (KNN)
Learning Type	Probabilistic, parametric	Non-parametric, instance-based
Speed	Very fast training and prediction	Slower due to distance computation
Data Requirement	Performs well with small datasets	Requires larger datasets for stable results
Interpretability	High (probability-based)	Moderate (depends on distance metric)
Sensitivity to Noise	Moderate; affected by assumption violations	High; influenced by noisy or overlapping points
Accuracy	Good baseline accuracy	Superior accuracy for nonlinear data
Scalability	Suitable for real-time embedded systems	Suitable for offline batch analysis
Best Use Case	Real-time, low-latency PD detection	High-precision offline diagnostic systems

Anusandhanvallari
Vol 2025, No.1
November 2025
ISSN 2229-3388

The comparative table clearly shows that GNB offers faster, lightweight classification suitable for on-site diagnostics, while KNN provides enhanced accuracy, particularly when high-dimensional nonlinear patterns are present in the PD dataset.

Trade-off Between Accuracy and Computational Cost

The trade-off between accuracy and computational cost is a key factor in selecting the optimal model for PD classification. GNB, with its simplified probabilistic framework, achieves near-real-time performance but slightly lower accuracy due to its feature independence assumption. KNN, conversely, provides higher precision but incurs higher computational overhead since it computes distances for every test instance. For online monitoring systems, where latency and resource constraints are critical, GNB's faster inference time makes it preferable. However, for offline analysis, such as laboratory diagnostics or maintenance planning, KNN's accuracy advantage outweighs its slower computation. Hybrid approaches—using GNB for quick detection followed by KNN refinement—could yield balanced performance.

Insight into Real-World Applicability

In real-world high-voltage systems, PD detection must balance accuracy, interpretability, and computational efficiency. The findings from this study indicate distinct application domains for both models:

- Online Monitoring: GNB is well-suited for embedded systems where rapid fault detection is necessary, such as transformer monitoring units or on-site PD testers. Its low processing requirement allows deployment on microcontrollers or FPGA-based setups.
- Offline Analysis: KNN, owing to its superior pattern recognition capability, is more appropriate for post-diagnostic data interpretation, laboratory research, and maintenance decision-making.
- Industrial Integration: Combining both models in a two-stage diagnostic framework—GNB for fast anomaly screening and KNN for refined classification—can provide reliable, scalable, and interpretable PD diagnostics.

The comparative evaluation of GNB and KNN demonstrates that both models are effective for PD source classification, each excelling in specific operational contexts. GNB provides computational efficiency and ease of implementation, while KNN achieves higher accuracy through flexible decision boundaries. The trade-off identified between accuracy and cost serves as a guideline for designing intelligent, resource-aware PD detection systems adaptable to both online and offline environments.

Conclusion and Future Scope

This study conducted a comprehensive comparative analysis of Gaussian Naive Bayes (GNB) and K-Nearest Neighbor (KNN) algorithms for identifying and classifying partial discharge (PD) sources in high-voltage insulation systems. Using an experimentally developed PD test bench, three primary discharge types—surface, internal, and corona discharges—were simulated to represent realistic insulation defects. The acquired PD signals underwent preprocessing, including noise filtering, normalization, and segmentation, followed by extraction of statistical, time-domain, and frequency-domain features to form the basis for machine learning classification. The results demonstrated that both models effectively distinguished between different PD sources. The KNN classifier achieved an accuracy of 96.8%, outperforming GNB, which attained 93.2%, due to KNN's superior ability to model nonlinear boundaries and handle overlapping feature distributions. However, GNB proved significantly faster and computationally efficient, offering real-time response capability. Misclassification was primarily observed between surface and internal PD patterns, attributed to feature similarity and correlated signal attributes.

The comparative findings highlight that GNB is more suitable for online PD monitoring systems, where rapid decision-making and lower computational demand are critical, while KNN is preferable for offline fault diagnosis, where higher precision and accuracy are prioritized. The research underscores the growing relevance of lightweight machine learning techniques in advancing intelligent condition monitoring, reducing maintenance costs, and enhancing power system reliability.

Future Scope

Future research can explore hybrid and ensemble learning approaches, integrating probabilistic and distance-based methods to achieve improved generalization and robustness. Incorporating deep learning architectures such as convolutional and recurrent neural networks may further automate feature extraction and enhance discrimination among complex PD patterns. Additionally, implementing edge computing and IoT-enabled systems for real-time PD detection can bridge the gap between laboratory models and practical field applications. Expanding the dataset with diverse insulation materials, voltage levels, and environmental conditions will strengthen model adaptability for industrial-scale deployment.

References

[1] Dalila, R.A.M., Turkben, A.K. Artificial intelligence based partial discharge detection using CNN and KNN to increase the quality of electrical insulation. Discov Computing 28, 101 (2025). https://doi.org/10.1007/s10791-025-09624-z

[2] Saravanabalaji Manian, Beena Stanislaus Arputharaj, Mohammad Imtiyaz Gulbarga, Subhav Singh, Deekshant Varshney, Mohankumar Subramanian, Abdulrajak Buradi, Vijayanandh Raja, N Sirisha, C Ahamed Saleel, Artificial Intelligence-Based Pipeline Leakage Detection in Water Distribution Networks for Liquid-Cooled Battery Thermal Management Systems, Desalination and Water Treatment, 2025, 101557, ISSN 1944-3986, https://doi.org/10.1016/j.dwt.2025.101557.

[3] Sebahattin Serhat Turgut, Aydzhan Myustedzhebov, Aberham Hailu Feyissa, Low-cost multispectral sensor reveals cold chain breaks, meat type, and storage time in chicken meat samples, Food Control, Volume 167, 2025, 110816, ISSN 0956-7135, https://doi.org/10.1016/j.foodcont.2024.110816

[4] Atmaja, B.T., Ihsannur, H., Suyanto et al. Lab-Scale Vibration Analysis Dataset and Baseline Methods for Machinery Fault Diagnosis with Machine Learning. J. Vib. Eng. Technol. 12, 1991–2001 (2024). https://doi.org/10.1007/s42417-023-00959-9

[5] Wojtecki, Ł., Bukowska, M., Iwaszenko, S., & Apel, D. B. (2024). Machine Learning-Based Classification of Rock Bursts in an Active Coal Mine Dominated by Non-Destructive Tremors. Applied Sciences, 14(12), 5209. https://doi.org/10.3390/app14125209

[6] Altamimi, A., Alarfaj, A.A., Umer, M. et al. An automated approach to predict diabetic patients using KNN imputation and effective data mining techniques. BMC Med Res Methodol 24, 221 (2024). https://doi.org/10.1186/s12874-024-02324-0

[7] Sekatane, P. M., & Bokoro, P. (2023). Partial Discharge Localization through k-NN and SVM. Energies, 16(21), 7430. https://doi.org/10.3390/en16217430

[8] Chen X, Aljrees T, Umer M, et al. Cervical cancer detection using K nearest neighbor imputer and stacked ensemble learningmodel. DIGITAL HEALTH. 2023; 9. doi:10.1177/20552076231203802 https://doi.org/10.1177/20552076231203802

[9] Magsi, A. H., Mohsan, S. A. H., Muhammad, G., & Abbasi, S. (2023). A Machine Learning-Based Interest Flooding Attack Detection System in Vehicular Named Data Networking. Electronics, 12(18), 3870. https://doi.org/10.3390/electronics12183870

[10] R. Das et al., "Mathematical Morphology-Aided Bidirectional Long Short-Term Memory Network-Based Partial Discharge Pulse Sequence Classification," in IEEE Transactions on Plasma Science, vol. 51, no. 12, pp. 3454-3461, Dec. 2023, doi: 10.1109/TPS.2023.3330891. https://doi.org/10.1109/TPS.2023.3330891

[11] Bilgi, E., Karakus, C.O. Machine learning-assisted prediction of the toxicity of silver nanoparticles: a meta-analysis. J Nanopart Res 25, 157 (2023). https://doi.org/10.1007/s11051-023-05806-2

[12] Al Zaidawi, N. Q. J. (2022). Partial discharge detection using convolutional neural network and k-nearest neighbor algorithm. https://hdl.handle.net/11363/4086

[13] Ramón A, Torres AM, Milara J, Cascón J, Blasco P, Mateo J. eXtreme Gradient Boosting-based method to classify patients with COVID-19. Journal of Investigative Medicine. 2022; 70(7):1472-1480. doi:10.1136/jim-2021-002278 https://doi.org/10.1136/jim-2021-002278

[14] K. Sit, A. K. Pradhan, S. Dalai and B. Chatterjee, "Machine Learning Based Wavelet Decomposition Aided Surface Contamination Prediction of Outdoor Silicone Rubber Insulators," 2022 IEEE Calcutta Conference (CALCON), Kolkata, India, 2022, pp. 332-336, doi: 10.1109/CALCON56258.2022.10060105.https://doi.org/10.1109/CALCON56258.2022.10060105

[15] Bechelli, S., & Delhommelle, J. (2022). Machine Learning and Deep Learning Algorithms for Skin Cancer Classification from Dermoscopic Images. Bioengineering, 9(3), 97. https://doi.org/10.3390/bioengineering9030097

[16] Aslam, M., Ye, D., Tariq, A., Asad, M., Hanif, M., Ndzi, D., Chelloug, S. A., Elaziz, M. A., Al-Qaness, M. A. A., & Jilani, S. F. (2022). Adaptive Machine Learning Based Distributed Denial-of-Services Attacks Detection and Mitigation System for SDN-Enabled IoT. Sensors, 22(7), 2697. https://doi.org/10.3390/s22072697

[17] Chaudhari, K., Mishra, A. K., Kothoke, P., & Chaudhari, Y. (2021). Examination and Determination of Partial Discharge Source using Gaussian Naive Bayes (GNB) Technique. Turkish Online Journal of Qualitative Inquiry, 12(7).

[18] Nag, A., Gerritsen, A., Doeppke, C., & Harman-Ware, A. E. (2021). Machine Learning-Based Classification of Lignocellulosic Biomass from Pyrolysis-Molecular Beam Mass Spectrometry Data. International Journal of Molecular Sciences, 22(8), 4107. https://doi.org/10.3390/ijms22084107

[19] Rapacz, S., Chołda, P., & Natkaniec, M. (2021). A Method for Fast Selection of Machine-Learning Classifiers for Spam Filtering. Electronics, 10(17), 2083. https://doi.org/10.3390/electronics10172083

[20] A. K. Das, D. Dey, B. Chatterjee and S. Dalai, "A Transfer Learning Approach to Sense the Degree of Surface Pollution for Metal Oxide Surge Arrester Employing Infrared Thermal Imaging," in IEEE Sensors Journal, vol. 21, no. 15, pp. 16961-16968, 1 Aug.1, 2021, doi: 10.1109/JSEN.2021.3079570. https://doi.org/10.1109/JSEN.2021.3079570

[21] Yuzhou Wang, Zhengfei Li, Huanxin Chen, Jianxin Zhang, Qian Liu, Junfeng Wu, Limei Shen, Research on diagnostic strategy for faults in VRF air conditioning system using hybrid data mining methods, Energy and Buildings, Volume 247, 2021, 111144, ISSN 0378-7788, https://doi.org/10.1016/j.enbuild.2021.11114

[22] Salleh, M. S. M. (2020). Partial Discharge Classification for XLPE Cable Joints Using K Nearest Neighbors Algorithm (Master's thesis, University of Malaya (Malaysia)).