IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 8, 2020, accepted December 18, 2020, date of publication December 24, 2020,
date of current version January 4, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3047125

Development of Hypergraph Based Improved Random Forest Algorithm for Partial Discharge Pattern Classification

SUGANYA GOVINDARAJAN[1], JORGE ALFREDO ARDILA-REY[2], (Member, IEEE), KANNAN KRITHIVASAN[1], JAYALALITHA SUBBAIAH[1], NIKHITH SANNIDHI[3], AND M. BALASUBRAMANIAN[1]

[1]Electrical and Electronics Engineering Department, SASTRA Deemed University, Thanjavur 613401, India
[2]Departamento de Ingeniería Eléctrica, Universidad Técnica Federico Santa María, Santiago de Chile 8940000, Chile
[3]Zoho Corporation, Chennai 600042, India

Corresponding author: Jorge Alfredo Ardila-Rey (jorge.ardila@usm.cl)

This work was supported in part by the Agencia Nacional de Investigación y Desarrollo through the Fondecyt Regular Project under Grant 1200055 and the Fondef Project under Grant ID19I10165, in part by the UTFSM through DST-FIST, PI_m_19_01 Project under Grant SR/FST/ETI-338/2013(C) (dated September 10, 2014) and Grant FST/MSI-107/2015(C), and in part by Tata Realty-IT City-SASTRA Srinivasa Ramanujam Research Cell of SASTRA Deemed University.

ABSTRACT Precise partial discharge (PD) detection is a key factor in anticipating insulation failures. The continuous efforts of researchers have led to the design of a variety of algorithms focusing on PD pattern classification. However, the trade-off between features taken up for classification and the detection rate continues to pose considerable challenges in terms of feature selection from acquired data, increased computing time, and so on. In this article, a Hypergraph (HG) based improved Random Forest (RF) algorithm by employing the Recursive Feature Elimination (RFE) algorithm (HG-RF-RFE), has been developed for PD source classification. HG representation of data is considered for obtaining statistical features, which turn out to be a subset of a set of all hyper edges called Hyper statistical features (Helly, Non-Helly, and Isolated hyper edges). HG-RF-RFE takes hyper statistical features and hyper edges as features for classification. The algorithm's efficiency is tested against noise-free PD data obtained from SASTRA High Voltage Laboratory, and large-sized noisy PD data obtained from High-Voltage Research and Test Laboratory at Universidad Técnica Federico Santa Maria (LIDAT). The robustness of the proposed algorithm is tested with both time and phase domain PD features using the Mathews Correlation Coefficient (MCC), harmonic mean-based feature Score (F1 Score) as evaluation metrics, and by k-fold validation technique. The proposed HG-RF-RFE achieved 98.8% accuracy with minimal features and significantly reduces computation time without compromising accuracy. It is worth mentioning that the HG-RF-RFE technique is superior to many state of the art algorithms in terms of feature elimination and classification accuracy.

INDEX TERMS Hypergraph, partial discharge, pattern classification, random forest, recursive feature elimination, statistical features.

I. INTRODUCTION

Partial Discharge (PD) measurement has been identified as a reliable insulation assessment diagnostic tool for high voltage equipment. In the dielectric material (solid, liquid, or gaseous), cavities, voids, cracks, and gaps are significant defects that lead to physical as well as chemical deterioration in insulated interfaces when subjected to high voltage stress. Whatever type of electrical equipment affected by PD can suffer from a series of severe insulation failures in the long term. The classification of PD patterns is an essential criterion for assessing and diagnosing the performance of the insulation systems, as it provides a significant index of discharge severity. The classification process aims to identify the defect that causes the discharge (surface discharge, corona, etc.) internally or externally. Since each defect has its typical degradation mechanism, in order to assess the quality of the insulation it is imperative to use this uniqueness to correlate

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaojie Ju.

the discharge patterns with the type of defect as part of the PD diagnosis.

PD diagnosis was usually done by visual inspection in the initial periods. Since the detection of PD is difficult, only experts with extensive personal experience are able to differentiate between various discharge phenomena and assess the severity of the fault. Although this way of identifying sources allows adequate recognition of the type of PD, in industrial environments the presence of external noise or multiple sources of PD can make it difficult to interpret the PRPD patterns. In the early 1990s, due to the availability of high-speed data processors and well-developed statistical techniques, tools for automated pattern recognition and machine-based expert systems were developed thus enabling more effective and efficient online monitoring of HV equipment. But now, in order to make the systems fully automated, Machine Learning algorithms are to be upgraded to meet the challenging issues in processing the different and varied data sets.

Researchers have focused on developing different techniques to classify PD patterns by extracting features considering these facts. Feature extraction exhibits the most discriminatory qualities in the raw data, enabling an accurate class detection of PD signals by the classifier. The appropriate features are selected to form the representative vector that allows the highest accuracy in recognizing patterns [1]. Numerous techniques for feature extraction, selection, and classification have been reported in the literature. Achieving low computational complexity and efficient performance with less computation time of the machine learning algorithms have become important issues for the machine learning community. Initial research studies have used a phase-resolved approach because it provides a good pattern recognition scheme [2], [3]. Various statistical measures using the phase window approaches were extensively used to achieve compact pulse signature characteristics of PD sources [4], [5]. Later, some researchers have also reported that time-solved pattern recognition methods have been applied successfully [6], [7]. Still, the statistical method of feature extraction offers a suitable mechanism for ageing characterization and provides an insight into the physics of the release process [8], [9].

Several research studies using traditional statistical functions as mathematical descriptors for PD patterns have been successfully conducted. However, redundant features cause more time consumption. Many machine learning techniques (Fuzzy sets, decision tree, K-nearest neighbor (K-NN), Hidden Markov Model (HMM), Support Vector Machine (SVM), etc.), Artificial Neural Networks (ANN) and its variants were successfully applied to PD classifications [10].

In [11], the authors reached a maximum of 92% classification accuracy by employing different versions of Probabilistic Neural Network (PNN) with (ψ, q, n) characteristics (phase/time of PD occurrence (ψ), PD pulse magnitude (q) and number of PD counts (n)) as input features. During the training phase, neural network-based studies showed better convergence. However, improvements in classification rate at reduced computing time remain a challenging task. The authors of [12] used the features of Principal Component Analysis (PCA) to classify PD sources using Support Vector Machine (SVM) and achieved a classification accuracy of 98%. However, the main drawback is that SVM training requires extensive matrix operations, which take time and are very slow. The various other PD pattern recognition techniques reported by the researchers are given in Table 1 [12]–[21].

The accuracy of PD classification techniques depends mainly on the features selected. Major shortcomings in the existing methods are (i) high detection rate with a greater number of features, and (ii) in most cases, synthetic noise was used to represent noisy data. Hence, the trade-off between either a high accuracy at the cost of longer computational times by using numerous features or an acceptably mediocre accuracy with faster computational time for continuous monitoring needs a re-look into features to check whether there exists inter-dependency (causing repetitive features). Consequently, obtaining minimal yet informative features which can classify PD pattern in less time with higher accuracy is the main research gap. This research work aims at developing one such classifier that can classify the PD sources with minimal features.

Most of earlier research works focused on design of machine learning algorithms with a wide variety of features in order to maximize the accuracy. The Random Forest algorithm was used in the recent research articles [22]–[24] to classify noisy and noise-free PD signals, and low-frequency band data in the frequency domain was considered as characteristic feature to classify the PD defects into three groups. The number of features selected by the research works shows that it is computationally more expensive. Nowadays, researchers are more interested in designing unsupervised learning-based classifiers to avoid complications in the extraction and selection of features since in Unsupervised learning, the algorithm itself can find patterns and associations between different data clusters. But the major drawback in deep learning is that it requires large amount of input data and extensive time to train the network [25].

Many works have shown that noisy input data can affect the performance of classifiers [26], [27]. Hence, the ubiquity of noise seems a significant issue in the practice of machine learning. While many machine learning techniques that can handle noisy data are reported in the literature, in most cases, the noise in the data is synthetic noise added with laboratory-measured data. In contrast, in this research work, the algorithm's efficiency is tested with PD data measured under two different laboratory conditions: shielded environmental conditions (less noise) and unshielded laboratory conditions (similar to PD measured at onsite) are considered [16], [28].

In this article, Hypergraph (HG) based Random Forest (HG-RF-RFE) has been proposed for classifying PD sources with minimal features by removing the redundant hyper edges. Using the Recursive Feature Elimination (RFE)

TABLE 1. Related works.

Techniques	Authors	Features Selected	Efficiency Percentage	Limitations
Support Vector Machine	Hao (2010) [13]	(ψ,q,n) characteristics (200 Features)	95.1%	SVM applied with wavelet co-efficient to classify PD sources, but generalization capability is poor
	Sarathi (2013) [12]	PCA Features (7 Features)	98.8%	SVM fails, to attain an accurate classification model from a small number of labelled examples
Neural Networks				
K-Nearest Neighbour	Norasage Pattanadech (2014) [14]	Frequency Ranges Statistical Features (8 Features)	86%	• For large data set classification problems, the structure of this model is complex
RHR-PNN	Venkatesh (2011) [15]	(ψ,q,n) characteristics (24 Features)	92%	• Computation time is high
A-NN	Raymond (2017) [16]	20 Statistical	97%	
Hidden Markov Model	Sathish (1993) [17]	Image Features	84%	Accuracy depends on the number of states and number of features
Fuzzy based Classifier	Abdel-Galil (2005) [18]	Pulse Characteristics (5 Features)	82.2 %	Computation cost and response time increase exponentially in proportion to the size of inputs.
Multi-Layer Perceptron	Hui Ma (2013) [19]	(ψ,q,n) characteristics (24 Features)	92%	Depends on the quality of the training and testing data
Radial Basis Function	NA Al-geelani (2012) [20]	Statistical Features (21 Features)	95%	Limited to Surface Discharge detection only
Bayesian Classifier	Karthikeyan (2005) [21]	(ψ,q,n) characteristics (108 Features)	96%	Interactions between features cannot be learned
ANFIS	Raymond (2017) [16]	Statistical Features (20 Features)	97%	Size and quality of training set and the features chosen to train affect the classification performance

process, the minimum informative hyper features for the Random Forest (RF) classifier are selected. It is necessary to select handcrafted features that are descriptive, discriminatory, and non-redundant to create an accurate and computationally efficient PD classifier. HG representation of raw PD data is chosen for this purpose since it is well suited for modelling complex relationships in the raw data. HG mathematical frameworks have been used in recent years as the basis for modelling networks, data structures, process scheduling, computations, and a variety of other systems where complex relations between the objects in the system play a dominant role.

In the proposed work, the Helly property of HG has been used to remove the redundant features from the measured PD data. After redundancy removal, statistical features in time and phase domain are extracted from the hyper features, and RF-RFE classifies the PD data with minimal statistical features by eliminating the unimportant features. Precisely, the HG introduction captures the pattern in a structured way through the neighborhood Hypergraph. The boosted novelty in the improved version of the classifier is the capture of the structural patterns through Helly hypergraph and the elimination of ill-structured and uninformative features through RFE. The performance of the proposed HG RF-RFE classifier over the existing classifiers' such as Naive Bayes (NB), Multi-Layer PNN (MLPNN), KNN, SVM (by using WEKA Tool), NGboost, XGboost, and LightGBM was evaluated with respect to the various performance metrics (Mathews Correlation Coefficient (MCC), Harmonic Mean based Feature Score (F1 Score), classifiers accuracy, precision, and recall), and by k-fold cross-validation technique [29]–[31].

II. PROBLEM FORMULATION

The main objective of this article is to design a recursive feature elimination algorithm and investigate the performance of RF classifiers with a reduced feature subset. A research study shows that, if features are extracted from signal data, the features may contain redundant attributes in the signals [10]. Hence, selecting the smallest, best feature subset for classification becomes a prime concern, while building the machine learning model would improve the performance and reduce the computation time. Although PCA is intended for feature reduction, its use poses a significant disadvantage since it does not consider class information [28]. In this article, HG has been applied first to extract Helly features. These Helly features help in removing the redundant features with less computing time and leads to an incremental accuracy at the end. The block diagram of the proposed feature elimination-based classifier is given in Fig. 1.

Hypergraph representation of data was implemented here for:

- Maintaining non-redundant hyper features
- Removing redundant features
- Extracting all hyper statistical features for classification.

Next, the RFE approach considers all statistical feature subsets and suggests the best feature subset for classification. This not only promises good performance but also shows the main features required for good classification. A set of statistical features in time and phase domain of different PD sources have been used as input feature vector for the design of RF classifier [8].

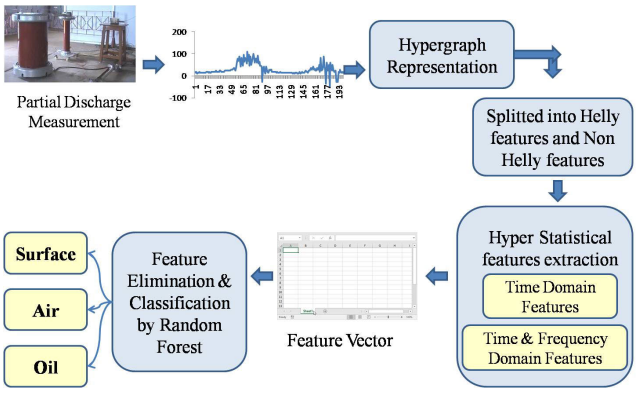

FIGURE 1. Block diagram of the proposed RF-RFE classification technique.

III. EXPERIMENTAL SETUP

PD signals measured at two different high voltage laboratories have been considered in this work to analyze the efficiency of the classifier. In case study I, PD signals were measured in a shielded environment at SASTRA High Voltage laboratory. In case study II, PD signals were measured in an unshielded environment that is prone to noise, in order to depict the PD signals measured in the industrial environment. Although the instrumentation used for each study case was completely different, the equipment, methods, and procedures used in each measurement were completely normalized [29].

A. CASE STUDY I: PD SIGNALS MEASURED AT SASTRA HIGH VOLTAGE LABORATORY

Three benchmark models related to single-source discharge patterns (i.e., surface discharge, air corona, and oil corona) were used to ensure the performance of the proposed algorithm for detailed studies. The arrangement of the PD test circuit and the diverse testing requirements comply with standard IEC 60270 [32].

Fig. 2 shows the test arrangement utilized to measure PD. The experiments were performed using a DTM-D Model- W. S. Test Systems system, consisting of the integrated oscilloscope (Tektronix-TDS 2002B), which has a tunable filter insert in the center-frequency range (0.6 MHz to 2.4 MHz) with a 9 kHz bandwidth for pulse acquisition of 2-5000 pC PD pulses. In this equipment the tunable filter acts as a variable selective filter for the selection of center frequency and to facilitate narrow band PD measurement with bandwidth of 9 kHz. The integrated oscilloscope displays PD pulses acquired from the analog output terminal, and the PD intensity is shown in terms of pico-coulomb (pC). HV Solution UK's PD Gold software was interfaced with the PD measuring system to obtain the PD patterns. The measurement test setup of a straight detection type was used for the test. The experimental analysis was carried out using 300 kV A; 300 kV; 50 Hz test transformer. A 1000 pF coupling capacitor and 1000 pF measurement impedance is attached to the test circuit in order to ensure enhanced pulse signal detection

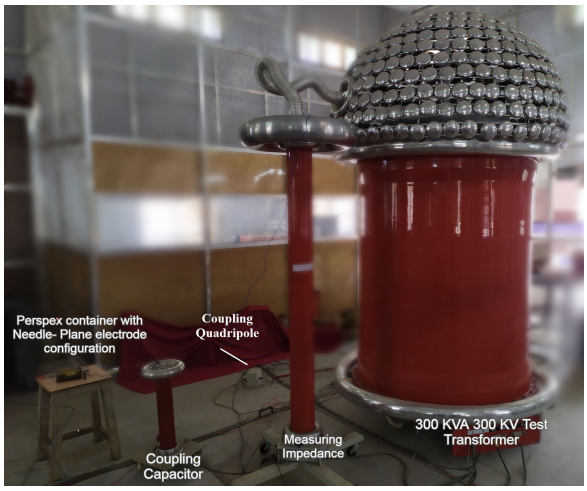

FIGURE 2. Case study I: Laboratory PD measurement setup with oil corona benchmark model.

and measurement. Here, the 1000 pF capacitor serves as a voltage divider to measure the output voltage from the test transformer.

The coupling quadrupole connected between the coupling capacitor and PD meter separates PD pulses superimposed on power frequency test voltage. With the help of a 5 mm thick dielectric sheet, one category of external discharge (surface discharge) was simulated, as shown in Fig. 3c. The air-corona discharge was replicated by an electrode of an apex angle 85 degrees attached to the high voltage (HV) bus shown in Fig. 3b. As shown in Fig. 3a, oil corona discharge was produced with a sharp point in transformer oil. The voltage across the needle-plane electrode was slowly increased until the PD was observed. The pulses were obtained at around 15-16 kV. The data is acquired at a sampling rate of 1 sample per 2.5 nanoseconds. PD data for each defect is recorded for 5 minutes, which provides 100 sinusoidal cycles with PD pulses on positive and negative half-cycles from the laboratory. The details are given in Table 2.

B. CASE STUDY II: PD SIGNALS MEASURED AT HIGH VOLTAGE LABORATORY OF UNIVERSIDAD TÉCNICA FEDERÍCO SANTA MARÍA

The experimental setup proposed in [28] for analyzing the separation of different types of PD sources is utilized here as case study II. The PD pulses are measured by indirect detection method given in standard IEC 60270 [32]. The experimental setup contains a high voltage source, a 1 nF capacitive divider of high voltage capacitor with measuring impedance (connected in series). The sensor used for the pulse's measurement was a high frequency current transformer (HFCT) with a bandwidth of 80 MHz. This inductive sensor has a ferromagnetic core and does not require galvanic contact with the electrical circuit during the measurement process. A NI-PXI-5124 digitizer with 200 MS/s of sampling rate, 12 bits of vertical resolution, and 150 MHz of bandwidth was programmed in LabVIEW to capture the PD pulses

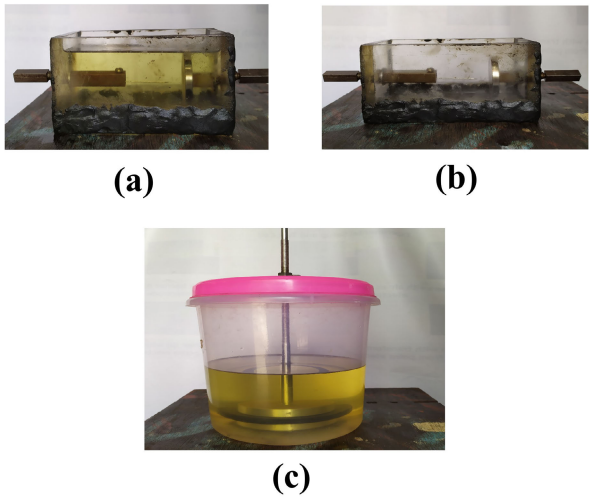

FIGURE 3. Laboratory benchmark model indicating a) Oil corona, b) Air corona and c) Surface discharge.

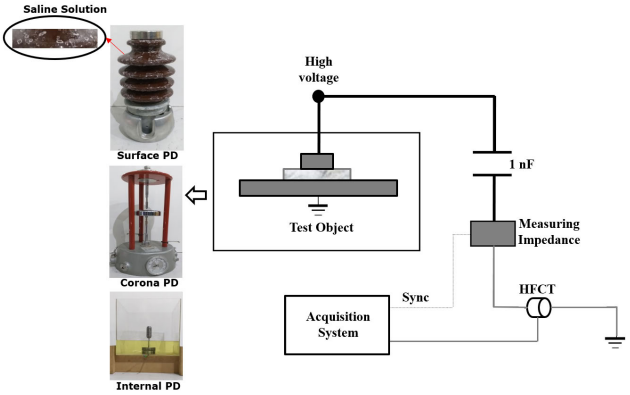

FIGURE 4. Case study II: Laboratory PD measurement setup.

during the experiment. Three PD sources, namely, corona defect, surface defect, and internal defect, were artificially created inside the unshielded environment. This ensures the electrical noise (impulsive noise of high-levels of amplitude and low-spectral variability) coupled with PD measurement. During the measurement process each signal was stored in time windows of 1μs or 4μs (200 or 800 samples), depending on the duration of the PD pulses. The experimental setup is depicted in Fig. 4. The measured PD pulses details are given in Table 2.

IV. CLASSIFICATION TECHNIQUE FOR PARTIAL DISCHARGE PATTERNS

In this article, the HG RF-RFE method is proposed to classify the PD patterns with minimum features. Also, well-known algorithms such as Neural Network (NN), SVM, NB, MLPNN, J48, OneR, NGBoost, XGBoost, and LightGBM have been used for PD source classification in order to compare them with the proposed method. As these classifiers were available in the Waikato Environment for Knowledge Analysis (WEKA) software, they have not been implemented separately [22].

TABLE 2. PD dataset acquired from benchmark laboratory models.

PD Category	PD Inception Voltage (kV)	No of PD Patterns
	Case Study I	
Air Corona	9 kV	101
Oil Corona	10.5 kV	102
Surface Discharge	7 kV – 8 kV	112
	Case Study II	
Corona Defect	5 kV	3000
Surface Defect	8.3 kV	3000
Internal Defect	9 kV	3000

In this section, the essential preliminaries of Hypergraph and Helly property to constitute Helly features and Random Forest and its recursive feature elimination strategies are discussed. The overall workflow is given in Fig. 5.

1) HYPERGRAPH PRELIMINARIES

Hypergraph (HG) is a new computing tool in clustered data avenues in the form of hyper edges which exhibit the relation among features in terms of both geometry and topology, mainly to reduce the computational burden of a chosen problem. For this reason, HG has been widely used by the research community for classification problems. Moreover, exploiting its properties to move towards their features, called "hyper features," requires machine intelligence to be built in any proposed technique. HG can represent the binary relation between vertices representing data points to n-array relations by including more than two vertices in an edge called hyper-edge. Mathematically, hypergraph can be defined as $H = \{X, E\}$, where $X = \{x_1, x_2, \cdots, x_n\}$ is the non-empty, finite set called vertices and $E = \{E_1, E_2, \cdots, E_m\}$ is the non-empty subset of X called hyper edges [33].

In many data science applications, when data points are represented as vertices, one can come across having pairwise relations, wherefrom an exciting and useful property (called Helly property) is observed. If all pairwise intersecting hyper edges have a common intersection, then that hypergraph is said to obey Helly property which represents the most prominent vertex among the neighborhood hyperedges. Consequently, there can be Helly (Fig. 6a) and Non-Helly (Fig. 6b), and Isolated hyper edges (Fig. 6c) can be observed once the data are represented in terms of hyper edges constituting a Hypergraph. A Machine learning algorithm learns Helly and Non-Helly and isolated hyper edges from the incidence matrices of the hypergraph [34], [35].

2) HYPERGRAPH BASED FEATURE SELECTION

Every feature selection method aims at a reduced number of features to achieve good classification accuracy when trained with any learning model. Textual, spectral, geometric, and statistical features are mainly used to train the learning model, but the error increases due to redundant features. Hence, HG is introduced for the identification of the optimal feature subset with minimal time complexity.

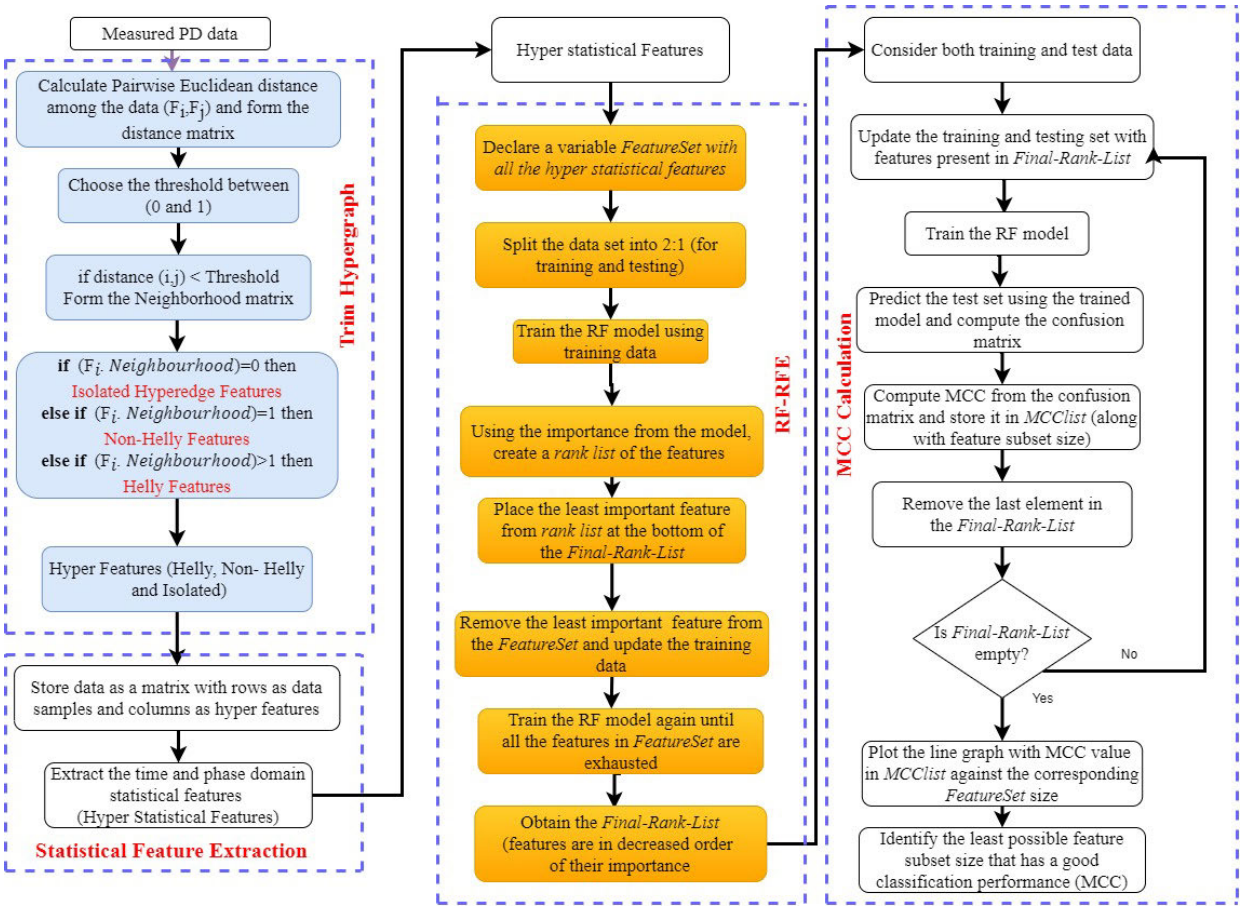

FIGURE 5. Flowchart of the proposed HG-RF-RFE.

The HG feature selection consists of two phases. In the first stage, hyper edges of the HG are obtained by establishing topological and geometrical connections between the features. The hyper edges and vertices correspond respectively to the data set samples and features.

In the second stage, the Helly property of the HG is applied to the intersecting hyper edges. In the further process, the features contained in the non-intersecting hyper edges are removed. The time complexity for identifying the ideal reduction is minimum because of HG's Helly property. If n is the number of the training sample, p is the number of features, ntrees is the number of trees (for methods based on various trees), the complexity of RF algorithm turns out to be $O(n2\sqrt{}pntrees)$ due to Breiman implementation. Since the number of generated hyper edges is not going beyond the number of training samples, there cannot be increased complexity. Hence the complexity of the proposed algorithm turns out to be the same as $O(n2\sqrt{}pntrees)$ [30].

3) BRIEF THEORY ON RANDOM FOREST BASED RECURSIVE FEATURE ELIMINATION

A collection of decision trees is generally a Random Forest. The motivation behind this idea is that learning made by a group of people would yield better results and allow the knowledge to be shared. Each decision tree predicts outcome, and RF collects the responses obtained from each tree and makes the overall decision. For classification, the class label classified by a majority of the decision trees is the final class given by the RF. In the case of regressions, the output of the RF is the mean of the outcomes produced by all the trees in the forest.

Bootstrap Aggregating (or bagging) is the process of splitting the samples randomly with replacement into B bags. Each bag is used to construct a decision tree, and therefore B decision trees will be built. In theory, one can show that when bootstrapping, only two-thirds of the data is allotted to a tree. The remaining one-third is termed as Out-Of-Bag (OOB). The advantage of random forest is that it uses this remaining one-third of data to find the error, known as the Out-Of-Bag error estimate.

If there are p predictors, it is generally recommended to have $\sqrt{p}$ feature candidates for each split in the case of a classification problem and p/3 feature candidates for that in the case of a regression problem. Nevertheless, it has to be considered that if the number of trees is very high, RF becomes computationally expensive.

The wrapper approach uses RF as a learning algorithm in this technique. The wrapper approach suggests the "most

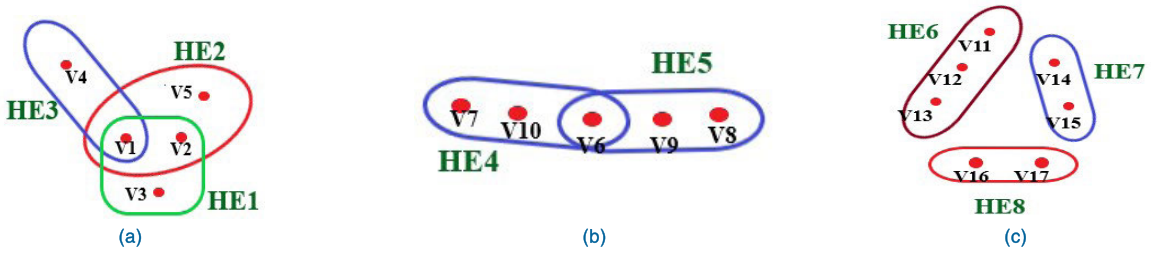

FIGURE 6. Hypergraph representation of data. (a) Intersecting hyper edges with common intersection (exploits Helly property) (b) Pairwise intersecting hyper edges without common intersection (lack of Helly property) (c) Isolated hyper edges.

important'' features for classification. The advantage of RF is that cross-validation is not required, as the Out-Of-Bag estimate can be used to get the parameter of the model. First, the optimal number of trees is obtained by training the RF model and choosing the number of trees that correspond to the lowest Out-Of-Bag error value. RF has been trained with an obtained number of trees. As discussed in the previous section, $\sqrt{p}$ number of predictors is tried for each split, where p is the number of predictors.

From the HG RF-RFE, a rank list of the feature candidates is obtained. The next task is to analyze the performance of each feature subset and choose the subsets that are having lower ranks. For each feature subset formed, RF has been trained and validated. The Matthews Correlation Coefficient (MCC) is computed from the confusion matrix. A line graph is plotted with every feature subset size against its corresponding MCC. Inferences are drawn from this graph and, by balancing the feature subset size with the performance, interesting feature subsets are selected for each setting (time domain and time-phase domain).

V. FEATURE EXTRACTION

A. SELECTION OF HYPER FEATURES

The Helly property of HG helps eliminating the redundant features in the data. Hyper edges are obtained after applying the HG algorithm, which reduces the structure of the PD data (Table 3). The threshold varies between 0.25 to 1.5, and the best accuracy was achieved when the threshold to select isolated hyper edges was 1. Redundant hyper edges are eliminated using the HG's trim property, as there may be an instance where these redundant hyper edges might be mistakenly predicted to apply to isolated hyper edges.

B. EXTRACTION OF STATISTICAL FEATURES

The number of features required in order to achieve successful classification depends on the distinctive quality of the selected features. During the last few decades, different approaches were taken to select characteristics in recognition of PD patterns. Techniques such as statistical approaches, pulse shape parameters, signal processing tools, image processing tools, and time-series approach, are widely reported in literature to extract features from the PD patterns [2].

Several statistical indexes have been proposed in the literature, and their classification accuracy is reported in Table 1.

TABLE 3. Hypergraph representation of data (case study II).

Original data size	2405*200
Data size after removing redundant features	1408*200

In general, two main distributions, pulse count and pulse height, can characterize PD data, and statistical features are extracted from these PD distributions [16]. Although there are many potential metrics available to identify PD, the authors are cautioned in selecting precise indicators that can uniquely and unambiguously classify the defects. Although some defects produce easily identifiable patterns, there may be a substantial overlap among them [26] and such situations may complicate accurate interpretation.

In this work, hyper statistical features related to time and phase domains are considered to classify the PD sources. Hyper statistical features are statistical features obtained from hyper-edge (HE) information that possess hyper relation property. Hence, redundant hyper edges with common information can be eliminated by the hypergraph algorithm. As a result, less hyper features are selected to represent the data.

Statistical and pulse shape characterizing features describing PD in the time domain are: F_1) Maximum Amplitude (A_{max}), F_2) Crest Factor (CF), F_3) Skewness (Sk), F_4) Kurtosis (Ku), F_5) Mean (μ), F_6) Variance (σ^2), F_7) Average Discharge Current (I_{Avg}), F_8) Quadratic Rate (QR), F_9) Discharge Power (P), and F_{10})I_{Avg}/QR. Statistical features describing PD in phase domain are: F_{11}) Phase angle of Maximum Amplitude (ϕ_{max}), F_{12}) Phase angle of Mean Amplitude (ϕ_{avg}), F_{13}) Kurtosis of max apparent charge phase distribution ($Ku_{\phi_{max}}$) F_{14}) Kurtosis of mean apparent charge phase distribution ($Ku_{\phi_{mean}}$), F_{15}) Skewness of max apparent charge phase distribution ($Sk_{\phi_{max}}$), F_{16}) Skewness of mean apparent charge phase distribution ($Sk_{\phi_{mean}}$). Their mathematical descriptions are as follow:

$$F_2 = CF = \frac{E_{max}}{E_{rms}} \tag{1}$$

$$F_3 = Skewness(S_k) = \frac{\sum_{i=1}^{N} (x_i - \mu)^3 f(x_i)}{\sigma^3 \sum_{i=1}^{N} f(x_i)} \tag{2}$$

$$F_5 = \text{Mean}\,(\mu) = \frac{\sum_{i=1}^{N} x_i f(x_i)}{\sum_{i=1}^{N} f(x_i)} \tag{3}$$

$$F_6 = \text{Variance}(\sigma^2) = \frac{\sum_{i=1}^{N} (x_i - \mu^2) f(x_i)}{\sum_{i=1}^{N} f(x_i)} \tag{4}$$

$$F_7 = I_{Avg} = \frac{1}{\Delta t} \sum_{i=1}^{N} |q_i| \tag{5}$$

$$F_8 = QR = \frac{1}{\Delta t} \sum_{i=1}^{N} q_i^2 \tag{6}$$

$$F_9 = P = \frac{1}{\Delta t} \sum_{i=1}^{N} q_i V_i \tag{7}$$

Here, N indicates the total number of discharges, Δt is the time interval, qi is the apparent charge, and V_i gives the magnitude of the voltage. To capture the Helly features, hypergraph data representation is done first, and Helly, non-Helly, and isolated hyper features are extracted. As mentioned in [2], hyper statistical features extracted both in time and phase domain (total 16) have been employed for classification through the HG-RF-RFE algorithm.

VI. RESULTS AND DISCUSSION

All samples are considered as an initial set of hyper edges and columns as features. The hyper features indicate selected hyper edges after the elimination of redundant and uninformative features by the RFE algorithm. To design a precise set of hyper edges, Helly hyper edges were also taken into account for determining the final set of hyper statistical features. The dataset was subjected to a chi-square test in order to test the hypothesis of independence among attributes. The hyper relation among the data is represented via hyper edges. Neighborhood relations between the hyper edges have been established through fixing a minimum threshold using the Euclidean distance algorithm. Chi-Square tests were performed with (m−1, n−1) degrees of freedom to determine whether the selected attributes are independent, where m and n represent the number of rows and columns, respectively [30]. From the dataset used in both case studies, it was observed that at least 65% of the attributes were found to be mutually dependent at a level of 5% significance.

In this section, the classification by the selected features is performed, and the results are compared with other classifiers by using k-fold cross-validation. The cross-validation method helps to control the over-fitting of the data. The k-fold cross-validation is the statistical practice that divides the sample data into k subsets. Out of k subsets, k−1 subsets are used for training the proposed system, and the remaining subset is used for testing the performance. In this article, a 3-fold, 5-fold, and 10-fold cross-validation method are used for analyses to authenticate the given samples.

Random Forest models are implemented by using the R packages available at the Comprehensive R Archive Network (CRAN) to classify PD sources. To verify effectiveness, the HG-RF-RFE classification performance is also estimated with different training/testing heuristics with training/testing ratios of 60:40, 70:30, and 80:20. Results pertaining to the classification of these sets ensure that change in the training-testing heuristics to ±5%, the difference in the detection rate is ±0:05%, and the results show that the proposed HG-RF-RFE can classify PD sources with a higher classification rate compared to other methods.

A. EVALUATION METRICS

The metrics used to quantify the performance of HG-RF-RFE are as follows:

1) MATTHEWS CORRELATION COEFFICIENT

The MCC is calculated from the confusion matrix and is one of the most reliable methods for assessing classifier's performance. The benefit of using MCC is that it captures the entire confusion matrix, thus giving a complete insight into the performance.

$$MCC = \frac{TP^*TN - FN^*FP}{\sqrt{(TP + FP)(FP + TN)(TN + FN)(FN + TP)}} \tag{8}$$

The value of MCC lies between −1 and 1. A value of 1 indicates excellent agreement, and the value of −1 indicates inferior prediction. MCC is used to compute the overall performance of the classifier.

2) ACCURACY

Accuracy (A_{value}) is simply a ratio between correct observation and total observations.

$$A_{Value} = \frac{TP + TN}{TP + FP + FN + TN} \tag{9}$$

3) PRECISION

Precision (P_{value}) is the ratio of correctly predicted positive observations of the total predicted positive observations.

$$P_{Value} = \frac{TP}{TP + FP} \tag{10}$$

4) RECALL

It is the ratio of correctly predicted positive observations to all observations in the actual class.

$$R_{Value} = \frac{TP}{TP + FN} \tag{11}$$

5) F1 SCORE

The F1 score is mathematically the harmonic mean of precision and recall that is given below in terms of the elements of the confusion matrix.

$$F_1SCORE = \frac{2TP}{2TP + FP + FN} \tag{12}$$

where TP, TN, FP, and FN are the True Positives, True Negatives, False Positives, and False Negatives, respectively, in the confusion matrix.

B. SELECTION OF INFORMATIVE FEATURES BY RF-RFE

The mean decrease in the Gini index is obtained from the trained model, and this metric tells how much Gini index decreases when a split of the node is made on a particular variable. The bigger the decrease, the more important is the variable. Based on the mean reduction in the Gini index, the variables of the current feature set are then sorted, and the least important variable is taken out, then it is placed at the bottom of the rank list, which is filled up in a bottom-up approach. The next task is to analyze the performance of each feature subset and choose the subsets that are having lower ranks. For each feature subset formed, RF has been trained and validated.

The mean of MCC is calculated from the confusion matrices obtained during cross-validation. The cost corresponding to the best MCC is chosen for training the RF using the combined data of training and validation sets. The final model obtained is used to predict the test set, and the confusion matrix is constructed. From this matrix, the final MCC is computed for this feature subset. The last element of the rank list is removed, and the entire process is repeated for the new subset of features. The algorithm ends when features in the rank list are exhausted

C. CASE STUDY I: CLASSIFICATION OF PD SOURCES MEASURED IN SHIELDED LABORATORY

In the first part of the experimental results, PD data measured in a shielded environment is considered, hence the noise level is very low. It is interesting to study the plots by observing how the classifier performs when features are removed one by one from the set. That is why the plot starts from the maximum subset size at the left, and size decreases when it goes to the right. From Fig.7a and 7b, it is evident that the MCC value is an almost straight line for all uninformative features (from feature 1 to 8). From feature 9, there is a sudden decrease in the MCC value, and the fall in MCC value indicates the presence of ''important'' feature for classification. The performance of the algorithm is analyzed first with the time domain features and later, with both time and phase domain features. A part of the database was utilized during the study and the models participating in the training were randomly selected. Fig. 7a and Fig. 7b show the HG-RF-RFE performance plot for time domain and time-phase domain features, respectively. According to decreasing MCC values, feature subset rank is assigned to the x-axis, and MCC values are assigned to the y-axis. The features obtained are plotted over the graph and the low rank features are encircled; those are the important informative features for classification. The sequence of features F2, F9, F7,……., F8 appearing in Fig.7a are the ranking of the features in the time domain and in Fig.7b phase domain ranking features are appearing in the sequence F9, F7, F2, ……., F16.

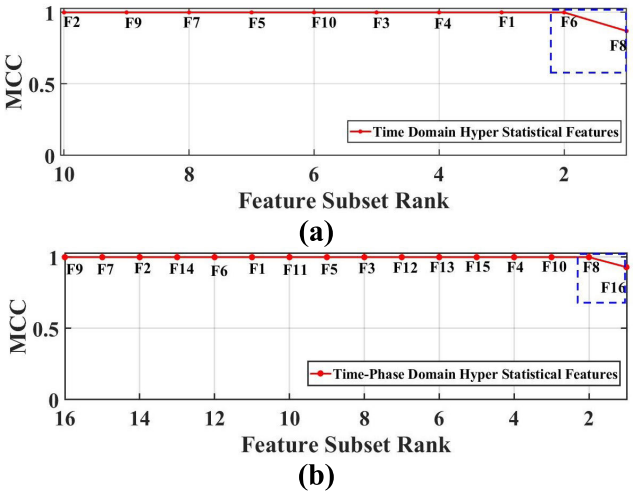

FIGURE 7. Case study I: MCC vs feature subset rank plot for a) time domain PD features, b) time-phase domain features.

From Fig. 7a, it is clear that there is almost a straight line up to 8 features indicating the presence of non-informative features. After the 9th feature (from left to right) there is a sudden decrease, which indicates the most relevant features for classification. Thus, in this case, features 6 and 8 (variance and QR) are chosen for classification. Likewise, in the time-phase feature subset, on a closer look, it is clear that the graph is similar to that of the time domain feature plot. i.e., the MCC value is equal to 1, for the 14 features, and there is a steep far after the subset size 15. Thus, in this case, the features 8 and 16 (variance and I/QR) are chosen for classification.

Since the quadratic rate assigns greater weight to the larger pulses, the QR of the discharge has been considered here as good indicator of the above three defects for the time domain features, and the feature I/QR is taken as the main indicator when time-phase domain features are considered. Along with QR, the variance has also been taken as a feature, which can tell about the widespread nature of the PD defects. Although the other statistical metrics suggested in [36] are good features, they are not precise indicators, and they require longer computation time. The efficiency of the HG-RF-RFE classifier over the existing classifiers is studied under various performance evaluation criteria shown in Table 4.

The hyper parameters that are used to tune the classifiers are given in Table 5. The main limitation in the SVM is the speed and size of the dataset in training and testing. For SVM, Radial Basis Function (RBF) kernel is used, which was actually found to be most effective for PD data. Moreover, when only two features are considered for classification, the recognition rate of the SVM is significantly reduced. The accuracy of NB is only 82% when the selected features are given as input to the NB classifier. The main reason for this low accuracy is that the features are dependent (HG helps to find the topological and geometrical relationship in the data). The classification accuracy of OneR is higher than SVM and NN. The proposed HG-RF-RFE is also compared with ensemble classifiers. The major drawback of ensemble type

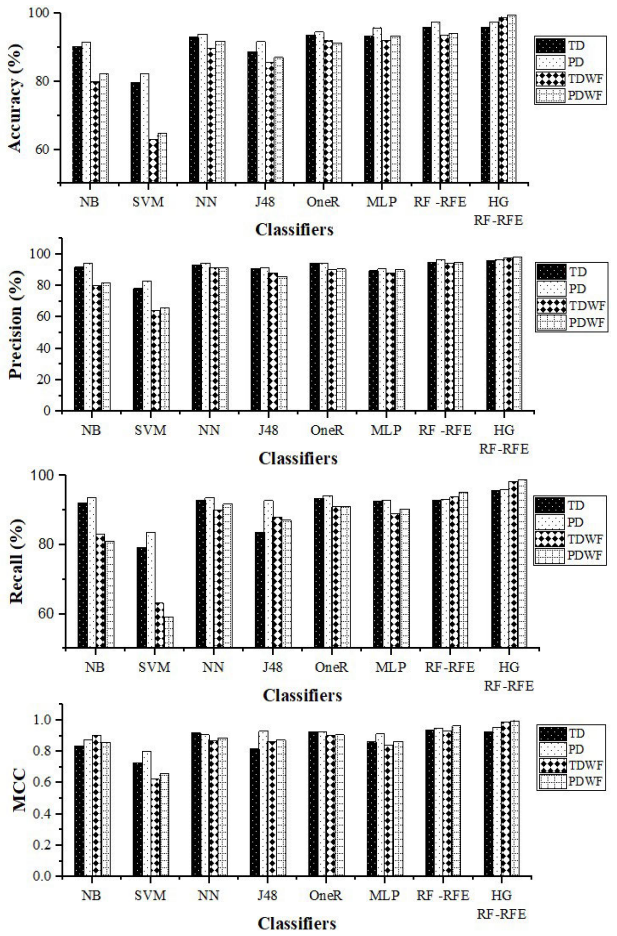

FIGURE 8. Performance analysis-accuracy, precision, recall, MCC values (top-bottom) for case study I PD data.

classifiers is that they take higher time in the training stage, and also that the recognition rate is very sensitive to data. Moreover, the accuracy of gradient boosting algorithms is nearer to the proposed HG- RF-RFE algorithm, but it requires more time to train, and it is harder to tune the parameters to classify effectively.

From the experimental analysis, it is evident that the performance of the HG-RF-RFE algorithm is better than the existing ones when the hyper statistical features are taken into account (Fig. 8). When the classification is done by all feature subsets, though the accuracy improvement is incremental, the performance of the classifier improves further when the selected hyper statistical features are considered. The variation in the accuracy level with the number of features has been observed in both time domain and time-phase domain feature subset for HG-RF-RFE classification.

D. CASE STUDY II CLASSIFICATION OF PD SOURCES MEASURED IN SHIELDED LABORATORY

In case study II, PD data measured from the unshielded laboratory is considered [28]. The case study I results showed that the proposed features could classify the PD sources

TABLE 4. Performance of different classifiers – case study I PD data.

Classifiers		Classifier Performance	
		Time Domain	Time – Phase Domain
NB	Accuracy	80.12	82.35
	Precision	80.32	82.14
	Recall	83.05	81.14
	F1 Score	0.81	0.82
	MCC	0.905	0.86
SVM	Accuracy	62.97	64.78
	Precision	64.01	66.17
	Recall	63.10	59.01
	F1 Score	0.64	0.62
	MCC	0.62	0.65
NN	Accuracy	90.01	91.89
	Precision	91.45	92.03
	Recall	90.14	91.81
	F1 Score	0.87	0.89
	MCC	0.87	0.89
J48	Accuracy	85.92	87.12
	Precision	88.3	86.13
	Recall	88.1	87.10
	F1 Score	0.82	0.87
	MCC	0.86	0.87
OneR	Accuracy	92.12	91.45
	Precision	90.29	91.09
	Recall	91.00	91.14
	F1 Score	0.90	0.92
	MCC	0.9	0.91
XGBoost	Accuracy	93.23	91.26
	Precision	93.12	91.31
	Recall	93.08	91.04
	F1 Score	0.92	0.91
	MCC	0.92	0.91
NGBoost	Accuracy	91.74	89.19
	Precision	91.24	89.05
	Recall	90.92	88.91
	F1 Score	0.89	0.87
	MCC	0.89	0.87
LightGBM	Accuracy	95.18	93.24
	Precision	95.26	93.12
	Recall	96.14	94.01
	F1 Score	0.95	0.93
	MCC	0.95	0.93
HG-RF-RFE	Accuracy	95.7	98.6
	Precision	96.2	98.3
	Recall	96.87	98.26
	F1 Score	0.96	0.98
	MCC	0.96	0.98

TD- Time Domain Features (with all hyper statistical features)
PD- Phase Domain Features (with all hyper statistical features)
TDWF- Time Domain Features (with features selected by HG-RF-RFE)
PDWF- Time-phase Domain Features (with features selected by HG-RF-RFE)

with higher accuracy than other state of the art classifiers. To prove the efficacy of the proposed classification method under noisy conditions, the hyper statistical features are alone taken into account. Such features indicate selective hyper edges after the elimination of redundant and uninformative features by the Trim Hypergraph. For this precise set of hyper edges, Helly hyper edges have also been taken into account to determine the final set of hyper statistical features for classification. Fig. 9a and Fig. 9b represent how the classifier performs when the hyper statistical features are removed one by one from the subset.

TABLE 5. Hyper parameters of classifiers.

Classifier	Hyper Parameter
SVM	Kernel: Radial Basis Function
Neural Networks	Learning rate: 0.97 number of layers: 3 Training cycles: 500
Random Forest	No of Trees: 158 Criterion: Gini index
Gradient Boosting	No of Trees: 130 Learning rate: 0.5

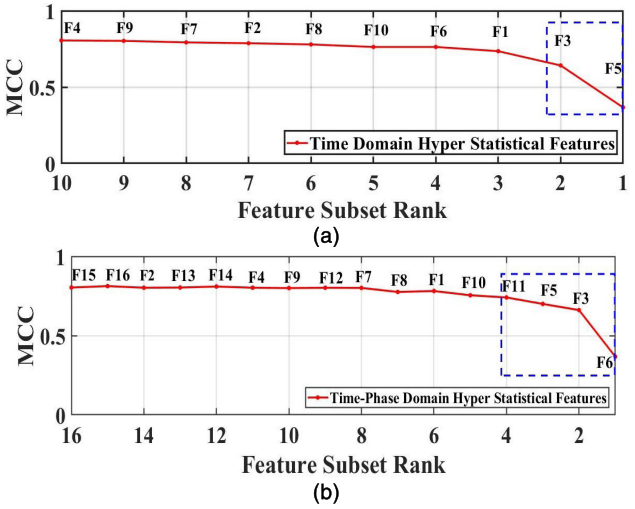

FIGURE 9. Case study II: MCC vs feature subset rank plot for a) time domain PD features b) time-phase domain features.

MCC was used to select the best feature subset, and this subset consists of hyper statistical features that are most informative and discriminative. The best feature subset is formed by grouping the hyper statistical features starting from the feature point where MCC begins to decrease and continues decreasing. The best feature subset (with time domain hyper statistical features alone) chosen as per the MCC score by the RFE algorithm, consists of four time domain hyper statistical features: variance, amplitude, kurtosis and mean (Fig. 9a). Compared to other PD classifiers, with these four time domain hyper statistical features, our proposed classifier (HG-RF-RFE) achieved the highest accuracy of 96.8%. This accuracy is achieved by forming a feature with the most informative and discriminative hyper statistical features from the time- domain. When time-phase domain features are considered, the accuracy of the algorithm is further increased to 99.8%.

The best-chosen feature subset consists of the following time domain and phase domain hyper statistical features: maximum amplitude, skewness, variance, I_{Avg}/QR, and phase angle of maximum amplitude (Fig. 9b), and the use of this feature subset as an input to HG-RF-RFE classifier improved its accuracy. The efficiency of the HG-RF-RFE classifier over the existing classifiers for case study II PD data is studied under various performance evaluation criteria shown

TABLE 6. Performance of different classifiers – case study II PD data.

Classifiers		Classifier Performance	
		Time Domain	Time – Phase Domain
NB	Accuracy	87.02	81.81
	Precision	87.23	82.59
	Recall	88.31	83.14
	F1 Score	0.87	0.83
	MCC	0.87	0.83
SVM	Accuracy	75.24	70.34
	Precision	75.12	70.57
	Recall	74.87	71.01
	F1 Score	0.74	0.70
	MCC	0.74	0.70
NN	Accuracy	92.47	91.45
	Precision	92.19	91.24
	Recall	93.08	92.21
	F1 Score	0.93	0.92
	MCC	0.93	0.92
J48	Accuracy	87.23	85.43
	Precision	87.08	85.20
	Recall	86.87	85.98
	F1 Score	0.87	0.85
	MCC	0.87	0.85
OneR	Accuracy	90.89	88.27
	Precision	91.26	89.09
	Recall	91.08	88.81
	F1 Score	0.91	0.88
	MCC	0.91	0.88
XGBoost	Accuracy	95.14	93.12
	Precision	95.47	92.14
	Recall	95.02	93.01
	F1 Score	0.95	0.93
	MCC	0.95	0.93
NGBoost	Accuracy	96.12	93.26
	Precision	96.24	93.24
	Recall	96.09	93.87
	F1 Score	0.96	0.93
	MCC	0.96	0.93
LightGBM	Accuracy	95.26	93.24
	Precision	95.64	93.64
	Recall	95.20	92.89
	F1 Score	0.95	0.93
	MCC	0.95	0.93
HG-RF-RFE	Accuracy	96.80	99.8
	Precision	96.15	99.41
	Recall	96.28	99.12
	F1 Score	0.96	0.99
	MCC	0.96	0.99

TD- Time Domain Features (with all hyper statistical features)
PD-Phase Domain Features (with all hyper statistical features)
TDWF- Time Domain Features (with features selected by HG-RF-RFE)
PDWF- Time-phase Domain Features (with features selected by HG-RF-RFE)

in Table 6. From the table, it is clear that the recognition rate of other state of the art algorithms is low when the selected time and phase domain hyper statistical features are given as input to the classifiers. Tuning of hyperparameters and hyper statistical features from both the time domain and phase domain, ensured the improved classifier accuracy. In this regard, the HG-RF-RFE classifier achieved a higher detection rate, and features were chosen by balancing the subset size with the performance.

The features selected for both case studies are shown in Table 7 for the training-testing ratio of (70%-30%), and the following conclusions have been summarized:

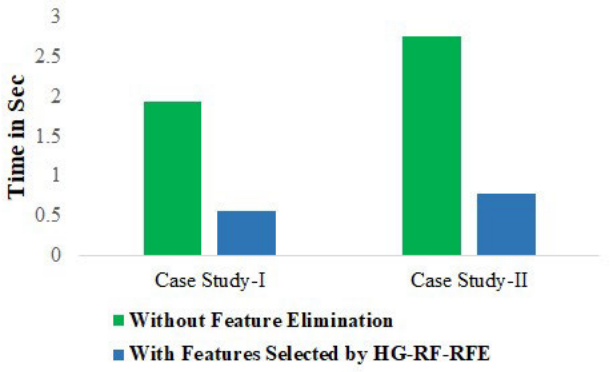

FIGURE 10. Computation time analysis.

TABLE 7. HG-RF-RFE classification performance for training testing ratio of (70%-30%).

Hyper Features	Selected Features	Classification Accuracy
	Case Study - I	
Time Domain	6 and 8	95.7 %
Time- Phase Domain	8 and 16	98.6 %
	Case Study - II	
Time Domain	1,3,5,6	96.8%
Time- Phase Domain	6,3,5,11	99.8%

- The proposed HG-RF-RFE algorithm has taken only hyper statistical non-redundant features to classify the PD defects into three different classes.
- As PD events are random, capturing and removing repetitive features is sufficient for efficient classification. Hence, non-Helly features are also included as HG-RF-RFE input.
- Some of the topologically significant redundant features are removed in the initial phase itself (HG phase). Statistical features are obtained only for the hyper-features obtained from HG phase outputs. These hyper-features are as precise as possible. Since isolated hyper edges are also made available at the beginning of the RF algorithm, no topologically significant feature is left.
- The size and nature of the data samples taken for training and testing are more sensitive to HG-RF-RFE.
- HG-RF-RFE is effective in classifying PD patterns and has the ability to eliminate redundant features. This is evident from Table 4 and Table 6 since, with only two features, a classification rate of 98.6% is accomplished in the case study I and 99.8% accuracy with four features in case study II.

What makes the RF random is that it selects a random subset of features for constructing a decision tree in the forest. Since the features are chosen and the bootstrapped data are random, this increases the diversity in the forest, making it robust and improving the performance of the classification. Due to this randomness, over-fitting is avoided, which is an advantage of this technique. One more benefit of this model is that it computes the importance of each feature and prefers the variables which are more important for classification.

TABLE 8. Average accuracy of classifiers using 3-fold, 5-fold and 10-fold cross validations on case study-ii PD data.

Classifier	Accuracy of 3-fold cross Validation	Accuracy of 5-fold cross Validation	Accuracy of 10-fold cross Validation
Naïve Bayes	82.8	84.3	81.81
SVM	68.7	72.4	70.4
NN	92.0	94.5	91.3
J48	86.9	82.3	85.4
OneR	88.3	90.2	88.27
XGBoost	94.4	93.6	93.12
NGBoost	92.29	94.01	93.26
LightGBM	93.01	94.98	93.24
HG-RF-RFE	99.1	98.7	99.8

The performance of the algorithm is compared with other existing classifiers, and the achieved detection accuracy proves the exhaustive capability of the HG-RF-RFE algorithm. Table 8 represents the accuracy of the different classifiers for the 3-fold, 5-fold and 10-fold cross-validations. The classifiers are trained with the time-phase domain features extracted from Case study II PD data. The minimal hyper statistical features selected by HG-RF-RFE (maximum amplitude, skewness, variance, I_{Avg}/QR and phase angle of maximum amplitude) are used to train other state of the art classifiers. It is evident, from the table, that the accuracy value of HG-RF-RFE is close to 100% across all chosen samples uniformly.

The consistency and convergence of the algorithm ensure the stability of the proposed method. From [37], it is evident that the increased number of features will make the system more complicated and time-consuming. Since HG-RF-RFE removes the redundant hyper edges and uninformative features, the time complexity and computation time of the algorithm are reduced. Fig. 10 shows the computation time of the HG-RF-RFE algorithm taken for classifying the PD sources when a reduced feature subset is considered for both case studies. In many practical applications, especially if online condition monitoring is necessary, this algorithm will select the vital feature to attain the best classification accuracy.

VII. CONCLUSION

Identification and classification of PD source is a challenging task. In the present work, the HG-RF-RFE method has been proposed to reliably classify PD sources with a minimum number of features. PD experiments have been carried out in the laboratory (shielded and unshielded environment condition) to simulate three different types of PD sources. HG is introduced mainly for two purposes: (i) to find out the hyper relation between the features, and (ii) to extract hyper statistical features based on time and phase domain. The reduced number of related features of PD signals (by taking time-phase domain features) is extracted for classification. As the redundancy and use of many features affect the performance

of the system, the RFE approach was chosen to find the best feature. Once the features are arranged in ranking using RFE, the ideal feature subset is selected with the help of MCC and employed to classify the PD source through Random Forest. It has been observed that, with selected features alone, the PD sources are classified with 99.8% accuracy after removing the uncorrelated features. The algorithm proposed in this article provides the potential hyper statistical features (statistical features obtained from HG representation of data) that can be used to classify the air corona, oil corona, and surface discharge defect types effectively. In the future, HG can be combined with parameterized hyper tree classifiers along with similar computational intelligent tools such as Grid Search, Generic Algorithm, Chi2 based feature selection, Mutual Information based feature selection and Extra Tree Classifiers to maintain this accuracy with less computational time using convex optimization techniques.

ACKNOWLEDGMENT

The authors would like to express their appreciation for creating the facilities at High Voltage Lab and Discrete Mathematics Lab of SASTRA, with which the experimental works were carried out.

1. The authors thank the Management of SASTRA University and gratefully acknowledge for the support extended to them in carrying out this research work.
2. The authors of this article gratefully acknowledge the comments and suggestions provided by the anonymous reviewers towards the improvement of this article.

REFERENCES

[1] F. H. Kreuger, E. Gulski, and A. Krivda, "Classification of partial discharges," *IEEE Trans. Electr. Insul.*, vol. 28, no. 6, pp. 917–931, Dec. 1993.

[2] N. C. Sahoo, M. M. A. Salama, and R. Bartnikas, "Trends in partial discharge pattern classification: A survey," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 12, no. 2, pp. 248–264, Apr. 2005.

[3] E. Gulski, "Digital analysis of partial discharges," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 2, no. 5, pp. 822–837, Oct. 1995.

[4] H.-G. Kranz, "Diagnosis of partial discharge signals using neural networks and minimum distance classification," *IEEE Trans. Electr. Insul.*, vol. 28, no. 6, pp. 1016–1024, Dec. 1993.

[5] B. Karthikeyan, S. Gopal, and S. Venkatesh, "Partial discharge pattern classification using composite versions of probabilistic neural network inference engine," *Expert Syst. Appl.*, vol. 34, no. 3, pp. 1938–1947, Apr. 2008.

[6] D. Evagorou, P. L. Lewin, V. Efthymiou, A. Kyprianou, G. E. Georghiou, A. Stavrou, and A. C. Metaxas, "Feature extraction of partial discharge signals using the wavelet packet transform and classification with a probabilistic neural network," *IET Sci., Meas. Technol.*, vol. 4, no. 3, pp. 177–192, May 2010.

[7] A. A. Mazroua, M. M. A. Salama, and R. Bartnikas, "PD pattern recognition with neural networks using the multilayer perceptron technique," *IEEE Trans. Electr. Insul.*, vol. 28, no. 6, pp. 1082–1089, Dec. 1993.

[8] E. Gulski and A. Krivda, "Neural networks as a tool for recognition of partial discharges," *IEEE Trans. Electr. Insul.*, vol. 28, no. 6, pp. 984–1001, Dec. 1993.

[9] A. A. Mazroua, R. Bartnikas, and M. M. A. Salama, "Discrimination between PD pulse shapes using different neural network paradigms," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 1, no. 6, pp. 1119–1131, Dec. 1994.

[10] W. J. K. Raymond, H. A. Illias, A. H. A. Bakar, and H. Mokhlis, "Partial discharge classifications: Review of recent progress," *Measurement*, vol. 68, pp. 164–181, May 2015.

[11] S. Venkatesh and S. Gopal, "Orthogonal least square center selection technique – a robust scheme for multiple source partial discharge pattern recognition using radial basis probabilistic neural network," *Expert Syst. Appl.*, vol. 38, no. 7, pp. 8978–8989, Jul. 2011.

[12] R. Sarathi, I. P. Merin Sheema, and R. Abirami, "Partial discharge source classification by support vector machine," in *Proc. IEEE 1st Int. Conf. Condition Assessment Techn. Electr. Syst. (CATCON)*, Dec. 2013, pp. 255–258.

[13] L. Hao and P. Lewin, "Partial discharge source discrimination using a support vector machine," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 17, no. 1, pp. 189–197, Feb. 2010.

[14] N. Pattanadech and P. Nimsanong, "Effect of noise signals on partial discharge classification models," in *Proc. IEEE Region Conf.*, Oct. 2014, pp. 1–5.

[15] S. Venkatesh and S. Gopal, "Robust heteroscedastic probabilistic neural network for multiple source partial discharge pattern recognition—Significance of outliers on classification capability," *Expert Syst. Appl.*, vol. 38, no. 9, pp. 11501–11514, Sep. 2011.

[16] W. Jee Keen Raymond, H. A. Illias, and A. H. Abu Bakar, "Classification of partial discharge measured under different levels of noise contamination," *PLoS ONE*, vol. 12, no. 1, Jan. 2017, Art. no. e0170111.

[17] L. Satish and B. I. Gururaj, "Use of hidden Markov models for partial discharge pattern classification," *IEEE Trans. Electr. Insul.*, vol. 28, no. 2, pp. 172–182, Apr. 1993.

[18] T. K. Abdel-Galil, R. M. Sharkawy, M. M. A. Salama, and R. Bartnikas, "Partial discharge pattern classification using the fuzzy decision tree approach," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 6, pp. 2258–2263, Dec. 2005.

[19] H. Ma, J. C. Chan, T. K. Saha, and C. Ekanayake, "Pattern recognition techniques and their applications for automatic classification of artificial partial discharge sources," *IEEE Trans. Dielectr. Electr. Insul.*, vol. 20, no. 2, pp. 468–478, Apr. 2013.

[20] N. A. Al-geelani, M. A. M. Piah, and R. Q. Shaddad, "Characterization of acoustic signals due to surface discharges on H.V. glass insulators using wavelet radial basis function neural networks," *Appl. Soft Comput.*, vol. 12, no. 4, pp. 1239–1246, Apr. 2012.

[21] B. Karthikeyan, S. Gopal, S. Venkatesh, and S. Saravanan, "PNN and its adaptive version-an ingenious approach to PD pattern classification compared with BPA Network," *J. Electr. Eng.*, vol. 57, no. 3, pp. 138–145, 2006.

[22] R. Hussein, K. B. Shaban, and A. H. El-Hag, "Robust feature extraction and classification of acoustic partial discharge signals corrupted with noise," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 3, pp. 405–413, Mar. 2017.

[23] R. Yao, J. Li, M. Hui, L. Bai, and Q. Wu, "Feature selection based on random forest for partial discharges characteristic set," *IEEE Access*, vol. 8, pp. 159151–159161, 2020.

[24] X. Peng, J. Li, G. Wang, Y. Wu, L. Li, Z. Li, A. Ahmed Bhatti, C. Zhou, D. M. Hepburn, A. J. Reid, M. D. Judd, and W. H. Siew, "Random forest based optimal feature selection for partial discharge pattern recognition in HV cables," *IEEE Trans. Power Del.*, vol. 34, no. 4, pp. 1715–1724, Aug. 2019.

[25] M. Karimi, M. Majidi, H. MirSaeedi, M. M. Arefi, and M. Oskuoee, "A novel application of deep belief networks in learning partial discharge patterns for classifying corona, surface, and internal discharges," *IEEE Trans. Ind. Electron.*, vol. 67, no. 4, pp. 3277–3287, Apr. 2020.

[26] H. Suzuki and T. Endoh, "Pattern recognition of partial discharge in XLPE cables using a neural network," in *Proc. 3rd Int. Conf. Properties Appl. Dielectr. Mater.*, 1992, pp. 43–46.

[27] J. Gao, Y. Zhu, and Y. Jia, "Pattern recognition of unknown partial discharge based on improved SVDD," *IET Sci., Meas. Technol.*, vol. 12, no. 7, pp. 907–916, Oct. 2018.

[28] J. Ardila-Rey, R. Albarracín, F. Álvarez, and A. Barrueto, "A validation of the spectral power clustering technique (SPCT) by using a rogowski coil in partial discharge measurements," *Sensors*, vol. 15, no. 10, pp. 25898–25918, Oct. 2015.

[29] L. Ali, C. Zhu, Z. Zhang, and Y. Liu, "Automated detection of Parkinson's disease based on multiple types of sustained phonations using linear discriminant analysis and genetically optimized neural network," *IEEE J. Transl. Eng. Health Med.*, vol. 7, pp. 1–10, 2019.

[30] L. Ali, C. Zhu, M. Zhou, and Y. Liu, "Early diagnosis of Parkinson's disease from multiple voice recordings by simultaneous sample and feature selection," *Expert Syst. Appl.*, vol. 137, pp. 22–28, Dec. 2019.

[31] A. Mas'ud, J. Ardila-Rey, R. Albarracín, and F. Muhammad-Sukki, "An ensemble-boosting algorithm for classifying partial discharge defects in electrical assets," *Machines*, vol. 5, no. 3, p. 18, Aug. 2017.
[32] *International Electrotechnical Commission (IEC) IEC 60270-High Voltage Test Techniques*, IEC, Geneva, Switzerland, 2000.
[33] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, "Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2406–2419, Jul. 2019.
[34] S. Nivedita, "Design and development of cloud service selection model using hypergraph based computational intelligence techniques," SASTRA Deemed University, Thanjavur, India, Tech. Rep., 2018.
[35] F. Luo, L. Zhang, B. Du, and L. Zhang, "Dimensionality reduction with enhanced hybrid-graph discriminant learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 8, pp. 5336–5353, Aug. 2020.
[36] M. Kunicki, A. Cichoá, and Å. Nagi, "Statistics based method for partial discharge identification in oil paper insulation systems," *Electr. Power Syst. Res.*, vol. 163, pp. 559–571, Oct. 2018.
[37] Z. M. Hira and D. F. Gillies, "A review of feature selection and feature extraction methods applied on microarray data," *Adv. Bioinf.*, vol. 2015, pp. 1–13, Jun. 2015.

SUGANYA GOVINDARAJAN received the M.Tech. degree in power systems in 2011. She has been in academia for the past seven years. She is currently a full-time Research Scholar with SASTRA Deemed University. Her research interest includes electrical insulation system and analysis especially for partial discharge pattern recognition and classification and denoising using computational hypergraph and computational intelligence techniques.

JORGE ALFREDO ARDILA-REY (Member, IEEE) was born in Santander, Colombia, in 1984. He received the B.Sc. degree in mechatronic engineering from the Universidad de Pamplona, Pamplona, Colombia, in 2007, the Specialist Officer degree in naval engineering from Escuela Naval Almirante Padilla, Cartagena, Colombia, in 2008, and the M.Sc. and Ph.D. degrees in electrical engineering from the Universidad Carlos III de Madrid (UC3M), in 2012 and 2014, respectively. He was an Automatic Control Engineer with ARC Almirante Padilla from 2008 to 2010. From 2010 to 2014, he was with the Department of Electrical Engineering, UC3M, and the High-Voltage Research and Tests Laboratory (LIDAT). He is currently a Professor with the Department of Electrical Engineering, Universidad Técnica Federico Santa María, Santiago, Chile. His research interests include partial discharges, insulation systems diagnosis, and instrumentation and measurement techniques for high-frequency currents.

KANNAN KRITHIVASAN received the bachelor's and master's degrees from the University of Madras, India, in 1980 and 1982, respectively, and the Ph.D. degree in mathematics from Alagappa University, Karaikudi, India, in 2000. He has been the Chair Professor (TIER-2) with Tata Realty-IT City-SASTRA Srinivasa Ramanujam Research cell in discrete mathematics for the past three years. He is currently a Professor with the Department of Mathematics, SARTRA Deemed University, Thanjavur, India. His specific areas of interest include combinatorial optimization, artificial neural networks, computational intelligence, and hypergraph-based image processing.

JAYALALITHA SUBBAIAH received the Ph.D. degree in high-voltage engineering from IIT Madras, India, in 2009. She is currently a Professor and the Dean of Research with SASTRA Deemed University. Her specific area of research includes high-voltage measurement and instrumentation techniques, analog circuits, and power system harmonics.

NIKHITH SANNIDHI received the bachelor's degree from the Department of Computer Science and Engineering, SASTRA Deemed University, in 2018. He is currently a Member of Technical Staff with Zoho Corporation, Chennai. His academic interests include algorithms, machine learning, and software engineering.

M. BALASUBRAMANIAN received the bachelor's degree in electrical and electronics engineering from Bharathidasan University in 2002 and the master's and Ph.D. degrees from SASTRA Deemed University in 2006 and 2017, respectively. His research interests include the condition assessment of power apparatus, aging characterization of insulation systems, and frequency response analysis for transformer fault diagnosis. He is currently a Faculty Member with the School of Electrical and Electronics Engineering, SASTRA Deemed University, Thanjavur, India. He is an Active Member of the IEEE and the IEEE Dielectrics and Electrical Insulation Society.