

# Use of Data-Driven Approaches for Defect Classification in Stator Winding Insulation

Emre Kantar<sup>1</sup>, Jaume M. Cascallo<sup>2</sup>, Torstein Grav Aakre<sup>1</sup>, Nina M. Thomsen<sup>1</sup>, and Espen Eberg<sup>1</sup>

<sup>1</sup> *Department of Electric Power Technology, SINTEF Energy Research*

<sup>2</sup> *Department of Electric Power Engineering, Norwegian University of Science and Technology*

## Abstract

Partial discharges (PD) in the high voltage insulation systems are both a symptom and cause of terminal and impending failures. The use of data-driven methods based on PD measurements will enable predictive strategies to replace traditional maintenance strategies. This paper employs machine learning-based classification models to identify and characterize PD signals originating from lab-made artificial defects in epoxy-mica material samples. Three different PD sources were studied: surface discharges in air, corona discharges, and discharges caused by internal cavities/delaminations. To generate high-quality datasets for the training, validation, and testing of classification models, Phase-Resolved PD (PRPD) data for each test object was obtained at room temperature under 50 Hz AC excitation at 10 % above the PD inception voltage (PDIV) of each sample. Relevant statistical and deterministic features were extracted for each observation and were labeled based on the defect type (supervised learning). Finally, the trained and validated machine learning (ML) models were used to identify PD sources in the service-aged stator winding insulation. Support vector machines (SVM), ensemble, and k-nearest neighbor (kNN) algorithms achieved significantly high accuracy ( $\geq 95\%$ ) of defect identification.

## 1. Introduction

Intermittent power generation causes stator winding insulation of hydrogenerators, which were designed for primarily continuous operation 50 years back, to experience damaging and frequent service failures, resulting in long downtimes, and costly repairs, thus, significant economic losses [1, 2]. Reliable and accurate condition monitoring of stator winding insulation has been widely performed using both off-line and on-line partial discharge (PD) measurements. PD measurements are a vital tool for assessing and monitoring the condition of power equipment. Different sources of PD have different effects on the insulation performance and thus reliability of the power apparatus. Therefore, identifying various PD sources at different locations is of great importance for the health assessment of stator winding insulation [3].

Data-driven methods employing artificial intelligence, such as machine learning (ML) algorithms, are now more feasible due to increasing computation power

and accessible tools, and a growing interest in such quantitative and predictive methods is expected. Developing such models for generator lifetime/condition qualification and estimation will accelerate the transition from traditional strategies toward predictive strategies, such as upgrading insulation components before they are estimated to fail [4].

Human experts usually perform a condition assessment of generator insulation with experience and qualitative judgment of data. The identification of PD sources (and their severity) is usually done using Phase-Resolved PD (PRPD) analysis, where each PD event is resolved into the apparent discharge magnitude ( $Q_a$ ), phase angle ( $\phi$ ), and the number of the PDs ( $n$ ). The use of ML-based data-driven models as a decision support tool can maximize the reliability and accuracy of the defect identification by unlocking hidden correlations.

ML-based PD diagnostics have been receiving increasing attention in the literature to handle the growing amount of data, reduce the human labor for feature engineering and tap the full potential of the data [5]. However, most of these studies employ similar fundamental features and are limited to specific test setups, whose experimental conditions do not explicitly state systematical details. These limitations inhibit comparative analysis and reproducibility. To further develop robust ML-based models, different statistical and deterministic PD features and due details for data acquisition and feature extraction techniques should be introduced.

The primary purpose of this study is to employ novel features (predictors) extracted from the obtained PRPD datasets and generate high-quality datasets for the training, validation, and testing of classification models based on various ML algorithms (classifiers). To that end, lab-made artificial defects are made to represent the most common defect types. Then, the qualified classification models (trained and tested) are used to predict possible defect types in the service-aged stator winding insulation.

## 2. Methodology

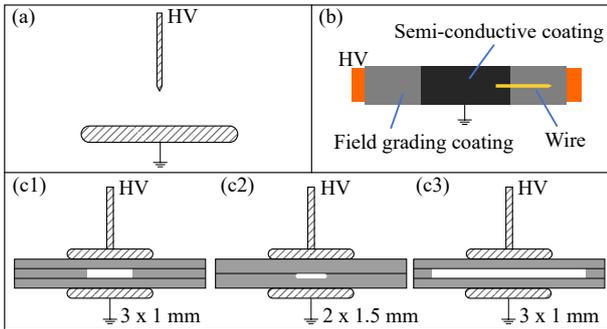
### 2.1. Test Objects

Test samples with known defects were made to emulate the most common discharge sources encountered in practice, which are classified into three main groups:

- i. corona discharges in the air (e.g., semiconductor-field grading paint (or tape) junction at the end windings),

- ii. surface discharges (e.g., slot discharges),
- iii. internal discharges (e.g., internal cavities and delaminations).

Fig. 1(a) illustrates the arrangement to induce corona discharges in an air gap (4 cm) between an energized piece of thin aluminum wire (with a tip diameter of 1 mm) and a flat ground electrode. To induce surface discharges, a similar thin wire was attached to the electrically grounded semiconductive coating and was extended towards the high voltage (HV) copper strands of a real hydrogenerator stator bar ( $\approx 2$  cm gap), as depicted in Fig. 1(b). In this work, stator bar is used interchangeably with stator winding insulation. Also, three service-aged stator bars (similar to the one shown in Fig. 1(b) without the wire) were used to predict PD sources associated with them.



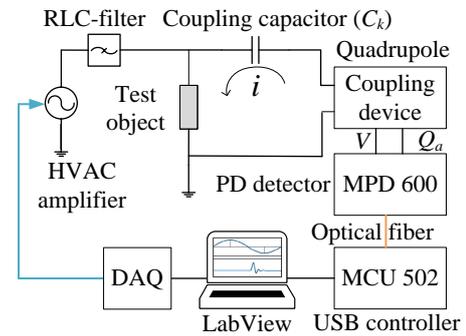
**Fig. 1** – Illustration of the lab-made samples as artificial PD sources. (a) Corona discharges (rod to plane). (b) Surface discharges. (c) Void discharges: c1. Center void: 10 mm void diameter and 1 mm void gap distance, c2. Delamination closer to the bottom electrode, and c3. Large void delamination near end-windings (voids/delaminations with unstressed walls).

We made laboratory objects from resin-rich mica/epoxy/glass-fiber tape to emulate internal discharges originating from voids and delaminations. The object dimensions were 100 mm  $\times$  100 mm  $\times$  3 mm. 1-mm of sheet/plate thickness was formed by stacking six tape layers half-overlapped. We cured the test objects at 160  $^{\circ}$ C for one hour under pressure. Metal spacers were used to create cylindrical voids during the curing process. Upon pressing cured plates together, a test object with a definite void dimension was formed.

Three different test objects with a total insulation thickness of 3 mm incorporating different void types were made, as exhibited in Fig. 1: (c1) 10 mm void diameter and 1 mm void gap distance, (c2) 5 mm void diameter and 0.5 mm void gap distance with round edges, and (c3) 40 mm void diameter and 1 mm void gap distance. The main difference between the voids in (c1) and (c3) is that the void in (c3) has electrically unstressed walls and represents large delaminations and voids. In (c2), the void was created only on one plate while the other plate was void-free, i.e., the void is slightly closer to the bottom electrode, representing an asymmetrical void in the insulation. The diameter of the HV electrode was 30 mm.

## 2.2. Test Setup

We generated high-quality PD datasets (free of crosstalk and noise above a selected threshold value of 1 pC) for the training, validation, and testing of various ML algorithms using a commercial PD acquisition unit (Omicron MPD 600) along with a 100-pF coupling capacitor. Fig. 2 illustrates the HV PD setup used for data generation, and *test object* stands for the lab-made samples with artificial defects. Desired voltage amplitude and frequency were set by a digital to analog converter (DAQ) and amplified to HV by an HV amplifier (TREK 20/20C–20 kV, 20 mA) in series with an RLC-low-pass filter with a cut-off frequency at 5 kHz. Measurements were performed according to IEC 60270, [6] where the amplitude spectrum was integrated with a center frequency of 250 kHz and a bandwidth of 300 kHz.



**Fig. 2** – Illustration of the experimental PD measurement setup.

## 2.3. Test Procedure and Data Acquisition

The test objects were subjected to voltage preconditioning before the data collection activity to preempt the so-called “memory effect”, and thus, the same test sample could be tested multiple times to achieve reliably extensive training and testing datasets. PD inception voltage (PDIV) of each test object had been experimentally determined in accordance with the definition in IEC 60270, and test objects were subjected to voltage preconditioning at 10 % above their corresponding PDIV at 50 Hz for 5 minutes before each test was performed (data collection activity). Given the stochastic nature of the PD phenomenon, generating data using the same test sample should be performed at least five times for acceptable statistical significance [7] as well as to generate an extensive and reliable dataset.

For this purpose, each sample was subjected to HV for 600 s at 50 Hz AC voltage. Each measurement (also referred to as “observation”) incorporated 1000 AC cycles at 50 Hz (20 s), allowing for up to a total of 30 separate data windows (30 observations) for the same test sample (30  $\times$  20 s = 600 s). Fig. 3 presents the steps followed for data and feature extraction, training, validation (fraction of training data used during training for tuning hyperparameters), testing, and prediction. MATLAB’s inherent functions and Classification Learner Toolbox were used for the entire data analysis and ML tasks. Critical remarks are listed below.

1. Employing the obtained PRPD data, statistical and PD quantities were extracted using 20 observations for each test sample out of 30 available observations: i.e., PD inception/extinction voltage, pulse repetition rate, phase angle, average discharge current, discharge power, quadratic rate, kurtosis, skewness. The generated parameters were labeled based on the defect type (class). The complete list of extracted features is shown in Table 1. The reasons for the selection of features are given in the next section.
2. The extracted features for each observation were labeled based on the artificial defect type (supervised learning). It should be highlighted that observations belonging to c1–c3 type defects were grouped/labeled together as "void discharge." The entire dataset was then split into training and testing sets using the hold-out method, where 80 % of the data was allocated for training, and the remaining 20 % of the data was used for testing. In addition, "k-fold cross-validation" was used on the training dataset where the dataset was arbitrarily divided into "k" groups. In this work, 5-fold cross-validation was used. One of the groups was used as the validation set, and the rest were used as the training set, as illustrated in Fig. 3. Then, the chosen model was trained on the training set and was evaluated on the test set. The procedure continued until each unique group was used as the test set.
3. Selected machine learning classifiers, such as support vector machines (SVM), k-nearest neighbor (kNN), ensemble, and decision tree, including their various subtypes, were trained by optimizing their hyperparameters. Then, predictions were made based on the defect type (label or class). Based on the training and test accuracies, the best models were chosen.
4. The qualified ML models were then used to predict/identify PD sources in service-aged (50 years) stator bars. Following the test procedure performed for the lab-made artificial defects, the same predictors/features were extracted from the obtained PRPD data of the stator winding insulation for each observation. The observations were then fed into the selected ML models, and possible PD sources were predicted based on the three training classes/labels.

The end-windings of these bars were removed due to the presence of asbestos, and then the semiconductive coating was removed at the ends, and field stress grading paint was applied by following the manufacturer's instructions. Thus, PD sources related to end-windings were eliminated. Also, the semi-conductive and field grading coating on the bars were checked against any abrasion or damage; only the undamaged bars were tested. Therefore, any PDs arising from the stator bars were expected to originate from internal discharges in the mainwall insulation represented by one or more void classes defined in c1–c3. However, the dataset was not labeled due to the uncertainty of unknown defects that might be present in the winding insulation. A detailed discussion of the predicted classes is given in the results.

## 2.4. Data Preprocessing and Feature Extraction

Firstly, the entire dataset was separately grouped for the positive and negative half-cycles as a reference to a sinusoidal AC voltage waveform to account for the asymmetrical PD events based on the defect type. The superscript (+) for the positive half cycle and (−) for the negative half cycle were used, respectively. Subsequently, numeric features (deterministic) such as number of charges, average, median, maximum (99th percentile employed to eliminate outliers) charge magnitudes, and discharge power were calculated for the positive and negative half-cycles (see Table 1) using the de-noised  $\phi - Q_a - n$  data exported from MPD 600.

Secondly, statistical post-processing of the  $\phi - Q_a - n$  data was performed in addition to the above-mentioned deterministic parameters. For the statistical analytics, PRPD data were characterized by using  $H_n(\phi)$ -phase distribution of the number of PDs and  $H_{qn}(\phi)$ -phase distribution of mean discharge amplitude. An example of such distributions extracted from the training database is shown in Fig. 4(a). Similarly, each feature was extracted separately for the positive and negative half-cycles using the corresponding  $H_n(\phi)$  and  $H_{qn}(\phi)$ , respectively.

We employed Weibull distribution, kurtosis (Ku), and skewness (Sk) analysis for the statistical analytics. The distribution of the obtained discharge amplitudes ( $Q_a$ ) follows an S-shape-like trajectory on the cumulative distribution function (cdf) plot, as presented in Fig. 4(b), indicating the existence of more than one active discharge mechanism [8]. The plot shows that a five-parameter mixed Weibull distribution (with two separate mechanisms) fitted the diffused discharge data well. A mixed Weibull distribution based on the sum rule is given by:

$$F(q) = p F_1(q) + (1 - p) F_2(q), \quad (1)$$

where  $q$  is the discharge magnitude ( $Q_a$  in our case),  $F_1(q)$  and  $F_2(q)$  are the cdfs of each discharge mechanism,  $p$  is the probability of occurrence of the subpopulation  $F_1(q)$  with  $0 \leq p \leq 1$  [8]. Below are five parameters;  $\alpha_1, \beta_1, \alpha_2, \beta_2$ , and  $p$ , shown explicitly:

$$F(q) = p \left\{ 1 - \exp \left[ - \left( \frac{q}{\alpha_1} \right)^{\beta_1} \right] \right\} + (1 - p) \left\{ 1 - \exp \left[ - \left( \frac{q}{\alpha_2} \right)^{\beta_2} \right] \right\}, \quad (2)$$

where  $\alpha$  is the scale parameter, and  $\beta$  is the shape parameter of each cdf. It is assumed that  $p = 0.5$ .

The shapes of  $H_n(\phi)$  and  $H_{qn}(\phi)$  have characteristic features for different PD sources [9]. Skewness and kurtosis have widely been used for the quantitative classification of PDs in artificial defects [9, 10] and hydro generators [11]. In relation to a normal distribution, skewness refers to the degree of asymmetry of a distribution: Sk is positive (negative) for a leftward (rightward) shift of a normal distribution. On the other hand, kurtosis stands for the degree of the sharpness of

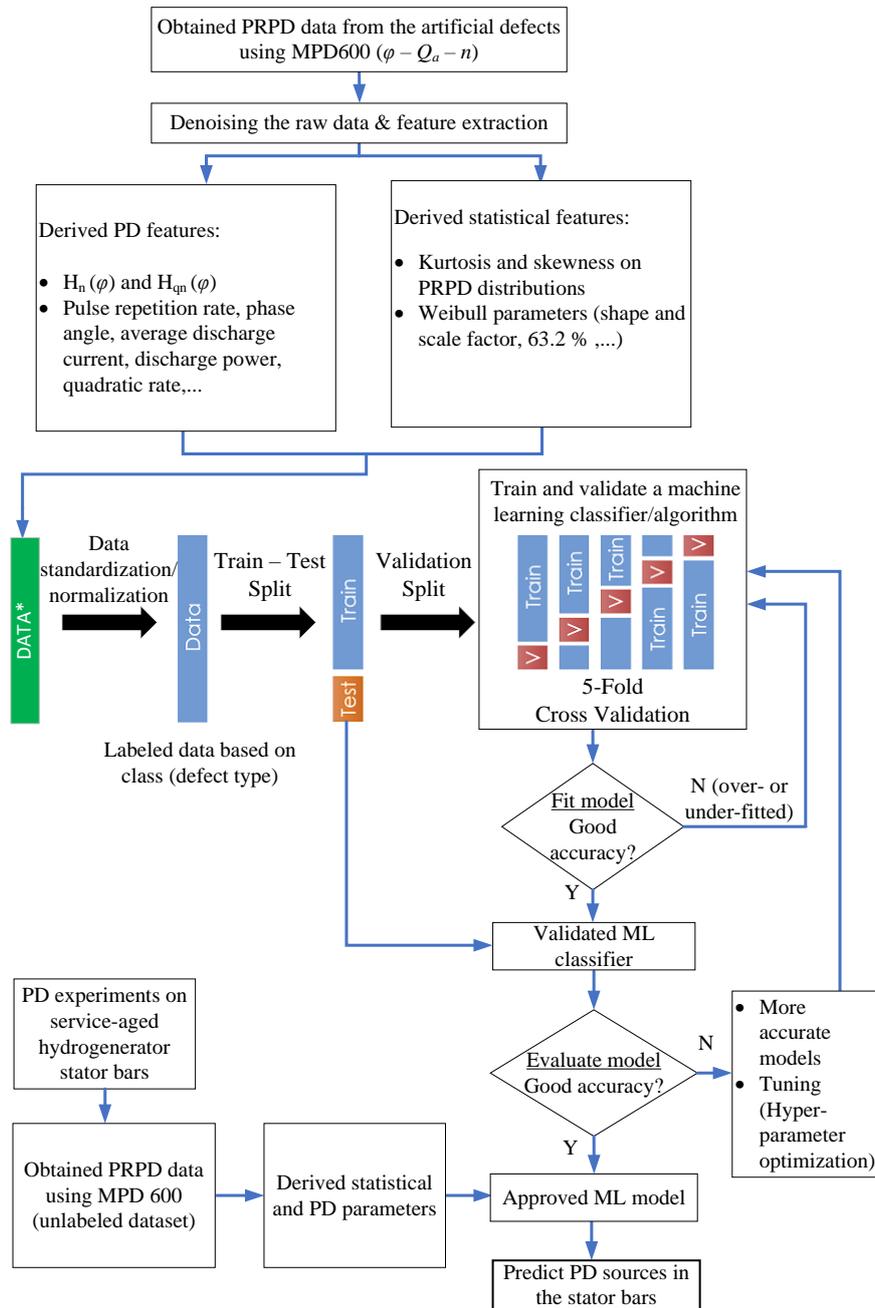


Fig. 3 – Flowchart showing the steps of data preparation, feature extraction, model training, testing, and tuning.

a distribution in relation to a normal distribution, and  $Ku$  is positive (negative) for a pointier (flatter) distribution compared to a normal distribution. Lastly, feature scaling was applied to the data to equalize their influence on the model. In this work, each feature column in the vector containing the entire observations ( $DATA^*$  as delineated in Fig. 3) was scaled to limit its range in the interval  $[0, 1]$ .

### 3. Results and Discussion

The obtained PRPD plots for 1000 AC cycles (50 Hz) for each defect type and winding insulation are shown in Fig. 5. The algorithms were then trained and tested based on the extracted features from the PRPD data to distinguish different PD sources under AC voltage. The input data consisted of the standardized 32 features

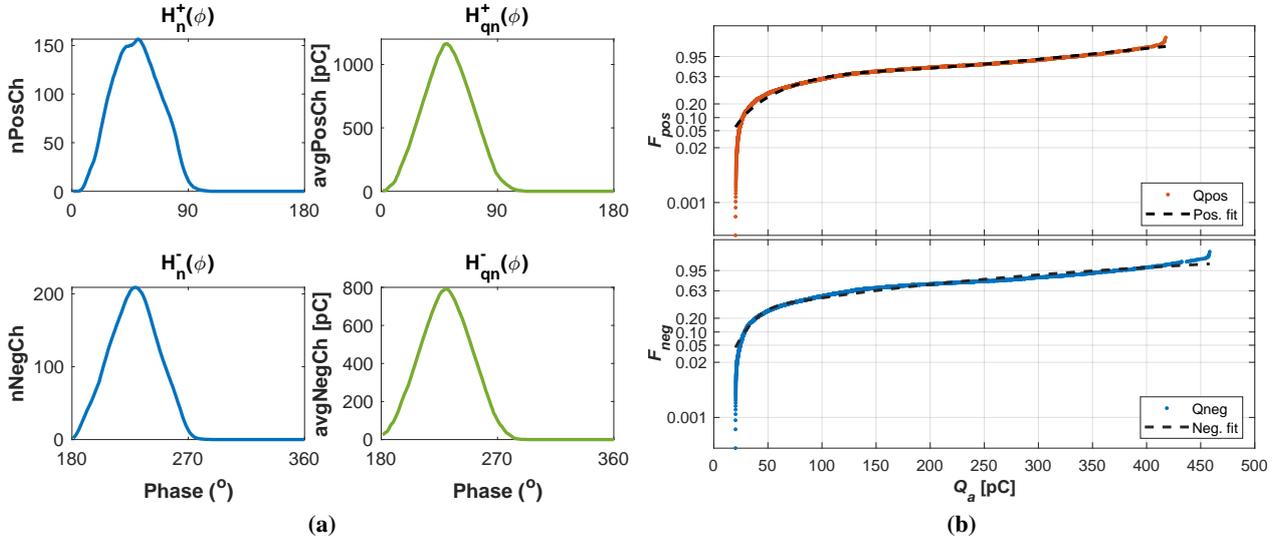
(Table 1). The predicted classes were rod to plane (*Rod2plane*), surface discharge (*surfaceDischarge*), and internal discharges (*Void*).

#### 3.1. Cluster Analysis and Feature Selection

A quick way to check if the classes can be grouped based on their extracted features is to visualize the data on a scatter plot and see if there are any obvious patterns or groups. For this purpose, chosen features were plotted against each other on a 3D scatter plot, as shown in Fig. 6. The selected features in the plot are closely related to most of the other features; hence similar patterns were observed among the classes. The scatter plot suggests that the extracted features can be useful to differentiate between the PD sources because the

**Table 1** – Features (predictors) extracted from the experimentally obtained PRPD datasets.

No.	Type	Feature (predictor)	Notation
1	Deterministic (PD)	Total number of pulses	nPulse
2	Deterministic (PD)	Number of positive charges	nPosCh
3	Deterministic (PD)	Average positive charge magnitude	avgPosCh
4	Deterministic (PD)	Median positive charge magnitude	medPosCh
5	Deterministic (PD)	Maximum positive charge magnitude	maxPosCh
6	Deterministic (PD)	Number of negative charges	nNegCh
7	Deterministic (PD)	Average negative charge magnitude	avgNegCh
8	Deterministic (PD)	Median negative charge magnitude	medNegCh
9	Deterministic (PD)	Maximum negative charge magnitude	maxNegCh
10	Deterministic (PD)	Number of pulses per cycle	ChRepRate
11	Deterministic (PD)	Average discharge current	avgCur
12	Deterministic (PD)	Total discharge magnitude per cycle	sumQ <sub>a</sub>
13	Deterministic (PD)	Average discharge power	avgP
14	Deterministic (PD)	Quadratic rate	D
15	Statistical (Weibull)	63.2% positive discharge magnitude	Qpos63
16	Statistical (Weibull)	63.2% negative discharge magnitude	Qneg63
17	Statistical (Weibull)	Scale parameter 1 (+)	ScalePos1
18	Statistical (Weibull)	Scale parameter 2 (+)	ScalePos2
19	Statistical (Weibull)	Shape parameter 1 (+)	ShapePos1
20	Statistical (Weibull)	Shape parameter 2 (+)	ShapePos2
21	Statistical (Weibull)	Scale parameter 1 (-)	ScaleNeg1
22	Statistical (Weibull)	Scale parameter 2 (-)	ScaleNeg2
23	Statistical (Weibull)	Shape parameter 1 (-)	ShapeNeg1
24	Statistical (Weibull)	Shape parameter 2 (-)	ShapeNeg2
25	Statistical (Kurtosis)	Phase distribution of number of discharges (+)	Ku (H <sub>n</sub> <sup>+</sup> )
26	Statistical (Kurtosis)	Phase distribution of mean discharge amplitude (+)	Ku (H <sub>qn</sub> <sup>+</sup> )
27	Statistical (Kurtosis)	Phase distribution of number of discharges (-)	Ku (H <sub>n</sub> <sup>-</sup> )
28	Statistical (Kurtosis)	Phase distribution of mean discharge amplitude (-)	Ku (H <sub>qn</sub> <sup>-</sup> )
29	Statistical (Skewness)	Phase distribution of number of discharges (+)	Sk (H <sub>n</sub> <sup>+</sup> )
30	Statistical (Skewness)	Phase distribution of mean discharge amplitude (+)	Sk (H <sub>qn</sub> <sup>+</sup> )
31	Statistical (Skewness)	Phase distribution of number of discharges (-)	Sk (H <sub>n</sub> <sup>-</sup> )
32	Statistical (Skewness)	Phase distribution of mean discharge amplitude (-)	Sk (H <sub>qn</sub> <sup>-</sup> )


**Fig. 4** – (a) Example of extracted  $H_n(\phi)$  and  $H_{qn}(\phi)$  distributions from the PRPD data of void discharge. (b) Fitted mixed Weibull cdf plots for positive and negative half-cycles retrieved from the PRPD data of void discharge.

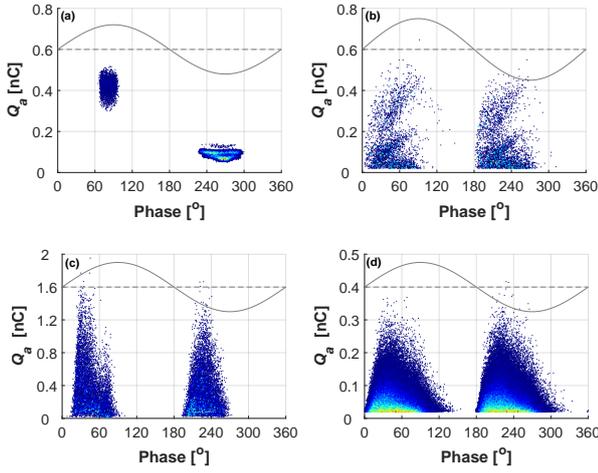
data points belonging to each class form distinguishable clusters. Also, using a box plot to compare classes after a statistical Weibull parameter, viz. ShapeNeg2 ( $\beta_2$  of the negative half cycle data) suggests discernible differences, as presented in Fig. 7.

### 3.2. Classification Analysis

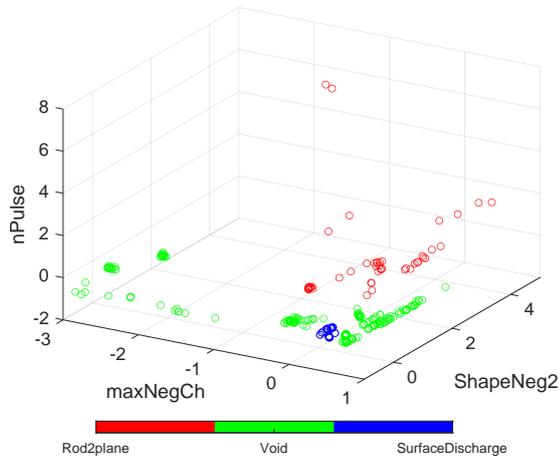
The performances of the trained and tested algorithms are usually interpreted using a confusion matrix, which depicts the amount of correctly classified instances in the

major diagonal and the wrong predictions in the minor diagonal. Also, ‘‘accuracy’’ is a useful parameter showing the overall performance and is defined as the value obtained by dividing the number of correct predictions by the number of total predictions.

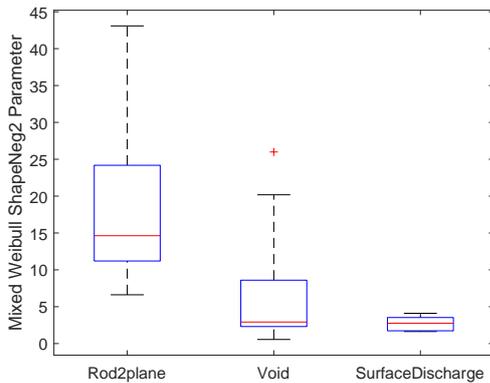
Table 2 shows a list of selected classifiers that were trained and validated along with their accuracies. Optimized versions for each classifier stand for the performed hyperparameter optimization (model tuning),



**Fig. 5** – Obtained PRPD from the lab-made samples. (a) Rod to plane. (b) Surface discharge. (c) Void discharge. (d) Discharges in the stator winding insulation.



**Fig. 6** – Scatter plot of selected features to visualize clusters.



**Fig. 7** – Box plot for the shape parameter (ShapeNeg2) vs. classes.

as illustrated in Fig. 3. As an example, Fig. 8 presents the confusion matrices for validation and test results of the tuned kNN-classifier. As depicted, the accuracy was 100 % both for the training (i.e., validation in training–222 observations) and testing (55 observations) data.

As the next step, selected validated ML models were used to predict/identify PD sources in the service-aged (50 years) stator winding insulation. The same data

**Table 2** – Summary of the accuracies of selected ML classifiers.  $k$ : no. of neighbors,  $\gamma$ : Kernel scale.

Classifier	Hyperparameters	Accuracy
Coarse kNN	$k=100$ , metric: euclidean, weight: equal	69.8 %
Cosine kNN	$k=10$ , metric: cosine, weight: equal	97.7 %
Fine kNN	$k=1$ , metric: euclidean, weight: equal	99.5 %
Optimized kNN	$k=1$ , metric: spearman, weight: squared inverse	100.0 %
Coarse Gaussian SVM	multiclass method: one-vs-one, $\gamma=23$	82.0 %
Fine Gaussian SVM	multiclass method: one-vs-one, $\gamma=1.4$	94.1 %
Linear SVM	multiclass method: one-vs-one, $\gamma$ : auto	99.1 %
Optimized SVM	Function: Cubic, multiclass method: one-vs-one, $\gamma=18$	100.0 %
Boosted tree Ensemble	method: AdaBoost, max. splits: 20, no. of learners: 30	69.8 %
Bagged tree Ensemble	method: bag, max. no. of splits: 221, no. of learners: 30	97.7 %
Optimized Ensemble	method: bag, max. no. of splits: 49, no. of learners: 11	100.0 %
Fine tree	max. splits: 100, split criterion: Gini's div. index	96.4 %
Optimized tree	max. no. of splits: 5, split criterion: max. dev. reduction	97.3 %

True Class	Optimizable kNN (Validation)			Optimizable kNN (Test)		
	Rod2plane	Void	surfaceDischarge	Rod2plane	Void	surfaceDischarge
Rod2plane	39			9		
Void		155			38	
surfaceDischarge			28			8

**Fig. 8** – Confusion matrix for the optimized kNN-classifier for training (validation) and test.

extraction methodology (de-noising, feature extraction, and data normalization/standardization) was followed before feeding the model with the data. PRPD data from three different service-aged winding insulation was acquired at their rated voltage (7.4 kV). 25 observations from each bar were obtained, amounting to a total of 75 observations where each measurement/observation incorporated 1000 AC cycles at 50 Hz (20 s). As previously mentioned, the stator bars did not contain any visible damage, nor did they have end-windings. Thus, a human expert would expect to observe PDs reminiscent of internal voids and delaminations. And based on the PRPD plots of the winding insulation, as shown in Fig. 5(d), they would label the PD source as internal discharge. However, we had not labeled the observations as "void" but as "unknown." Table 3 shows the predicted PD classes in the winding insulation by the selected models. The tuned SVM model classified 74/75 observations as "void," which was the expected response. Only one observation seems to be misclassified. The

optimized ensemble model yielded similar results, while the rest of the models predicted fewer voids.

As a caveat, ideally, a well-defined PD source in a different experimental setting would have been more assuring to test the model's accuracy, even though the initial results seem to be promising and indicate that the identification of different PD sources within a test object is viable. In future work, further tests will be performed on winding insulations with other types of known defects to endorse the models' accuracies more reliably. Then, the selected models will be tested with the unseen labeled experimental data representing each PD class. Moreover, not all the 32 features used in this work are likely to have the same importance for the trained models because some of them are close descendants of each other, and some may have random values that might mislead the training procedure and cause overfitting problems. After further endorsing the models using labeled unseen data, feature selection methods (i.a., neighborhood component analysis, MRMR, Chi2, ANOVA), dimensionality reduction techniques (i.a., principle component analysis, multidimensional scaling), and seeking different features will be performed to see if the models can be represented with fewer features without compromise in the overall accuracy. An overview of various qualified models will also be given based on the selected features. All pertinent ML classifiers, including neural networks, will be tested.

Last but not least, new classes to represent other types of PD sources will be included as well as defining different types of voids studied in this work as separate classes, and then the classifiers will be tested under those circumstances.

**Table 3** – Predicted PD sources in the stator bars (25 observations from each of three winding insulation).

Model	Rod2plane	Void	surfaceDischarge
Optimized SVM	1	74	0
Optimized Ensemble	3	72	0
Optimized kNN	7	68	0
Optimized Tree	49	24	2
Boosted tree Ensemble	49	13	13

## 4. Conclusions

The main aim of this work was to test out several ML-based models to classify different PD sources accurately, investigate the benefit of analytics for condition monitoring of HV insulation, and address challenges and knowledge needs for future studies. Lab-made artificial defects were used to generate  $\phi - Q_a - n$  datasets from which novel features were extracted. The generation of high-quality training datasets was a success, and they will be extended to incorporate more PD sources, especially those arising in stator winding insulation. Several ML classifiers were trained based on the extracted features, and validation accuracies thereof were benchmarked. Several types of SVM, ensembles and kNN models achieved significantly high accuracies

of PD defect identification ( $\geq 95\%$ ). Then, the selected classification models were used to predict possible defect types in service-aged winding insulations. The chosen models predicted the dominating PD source in the stator winding insulation to be of void discharge type, agreeing with the initial expectation. However, further tests should be performed on winding insulations with other known defects to more reliably verify the models' accuracy.

This study has laid the groundwork for the future investigation of feature selection and reduction, as well as the introduction of novel statistical features that are fingerprints of PD data. Finding the right features that adequately define and classify different PD sources is still ongoing for further improvement. In particular, five parameter Weibull fit approach yielded promising results and deserves further attention to identify simultaneously active PD mechanisms, e.g., slot discharges and internal discharges in the winding insulation.

## Acknowledgment

This work is funded by Digital Twin Laboratory for Hydropower Project through HydroCen FME Centre of Excellence (<https://www.ntnu.no/hydrocen/twinlab>).

## References

- [1] CIGRE WG A1.10, "TB 392 Survey of Hydrogenerator Failures," *e-cigre*, 2009.
- [2] C. Sumereder, "Statistical lifetime of hydro generators and failure analysis," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 15, no. 3, pp. 678–685, Jun. 2008.
- [3] M. Istad, M. Runde, and A. Nysveen, "A Review of Results From Thermal Cycling Tests of Hydrogenerator Stator Windings," *IEEE Trans. Energy Convers.*, vol. 26, no. 3, pp. 890–903, Sep. 2011.
- [4] M. Levesque, N. Amyot, C. Hudon, M. Bélec, and O. Blancke, "Improvement of a Hydrogenerator Prognostic Model by using Partial Discharge Measurement Analysis," Aug. 2018.
- [5] S. Lu, H. Chai, A. Sahoo, and B. T. Phung, "Condition Monitoring Based on Partial Discharge Diagnostics Using Machine Learning Methods: A Comprehensive State-of-the-Art Review," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 27, no. 6, pp. 1861–1888, Dec. 2020.
- [6] International Electrotechnical Commission, "IEC 60270:2000+AMD1:2015 CSV — High-voltage test techniques - Partial discharge measurements," Tech. Rep., 2015.
- [7] "IEC/IEEE Guide for the Statistical Analysis of Electrical Insulation Breakdown Data (Adoption of IEEE Std 930-2004)," *IEC 62539 First Edition 2007-07 IEEE 930*, pp. 1–53, 2007.
- [8] M. Cacciari, A. Contin, and G. Montanari, "Use of a Mixed-Weibull Distribution for the Identification of PD Phenomena [Rotating Machines]," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 2, no. 4, pp. 614–627, Aug. 1995.
- [9] T. Tanaka and T. Okamoto, "Analysis of q-n and  $\varphi$ -q characteristics of partial discharge in several electrode systems," in *1980 IEEE International Conference on Electrical Insulation*. IEEE, 1980, pp. 190–193.
- [10] E. Gulski, "Computer-aided measurement of partial discharges in HV equipment," *IEEE Transactions on Electrical Insulation*, vol. 28, no. 6, pp. 969–983, 1993.
- [11] E. Eberg, T. G. Aakre, G. Berg, and S. Hvidsten, "Comparison of Offline VLF PD Measurements and Online PD Measurements on a 50-Year-Old Hydrogenerator Stator in Norway," in *2018 IEEE Electrical Insulation Conference (EIC)*, Jun. 2018, pp. 542–546.