Fast Algorithm for Contactless Partial Discharge Detection on Remote Gateway Device

Tomáš Martinovič and Jan Fulneček

Abstract—**Detection of the high impedance fault caused by vegetation is one of the biggest challenges brought by the usage of covered conductors in medium voltage overhead power lines. One of the accompanying events of long-term contact of vegetation with the XLPE insulation is partial discharge which damages the insulation. Current systems for the detection of partial discharge have two major problems: the price and difficult installation. In this paper, an approach for the detection of the partial discharge from the data collected by the antenna is described. This approach is focused on the small computational demand and low false positive rate. Thanks to the small computational requirements, it can be run on the remote gateway devices which are collecting the data from the antenna. It is composed of four steps: outlier detection, outlier clustering, feature extraction, and classification. It is shown that this approach greatly improves the detection rate and lowers false positives compared to the previous algorithm used for partial discharge detection based on the data from an antenna, making it fit to use in the production environment.**

Index Terms—**Time series, partial discharge, pattern classification, outlier detection, wireless, high impedance fault, covered conductor.**

I. INTRODUCTION

MEDIUM voltage (MV) overhead power lines are usually equipped with regular AlFe conductors, without any external insulation. In a forested area, phase-to-ground and phase-to-phase faults often occur on such power lines because of the surrounding vegetation [1]. To eliminate such faults, AlFe conductors are being replaced with covered conductors (CC). CC consists of an aluminum core, which is covered by a thin layer of XLPE insulation material. The biggest disadvantage of CC lies in the problematic detection of high impedance fault (HIF), caused by vegetation. Because of the additional XLPE insulation, the fault current is extremely low, so it cannot be detected by standard protection relays. If a tree or branch stays in direct long-term contact with CC, partial discharges (PD) may appear on the surface of the XLPE insulation. PDs are causing slow degradation of CC, resulting in an insulation failure. In extreme situations, CC can be damaged after several hours of PD activity. However, usually it takes several days before severe insulation damage occurs. Early detection of PD activity can prevent the damage of CC and energy transmit interruption.

PD activity is always accompanying phenomena of HIF. Therefore, the presence of PD can be used as an indication of HIF occurrence. There are several ways how to acquire PD patterns from power lines with CC. Most of the methods uses galvanic methods for the acquisition of high frequency components of voltage or current signals. To obtain it, the measuring equipment had to be connected to the examined power lines via sensors: a capacitive divider (for voltage signals) or Rogowski coil (current signal). These sensors must be suitable for MV usage, which makes them expensive. This is why there are just a few online PD detectors for overhead power lines commercially available on the market-high cost prevent detectors to become widely spread. Lu *et al.* [2] provided a comprehensive summary of different algorithms used for the detection of partial discharge in various applications.

II. PROBLEM DEFINITION

To decrease price of the detector as much as possible, it was decided to replace the common galvanic sensors by a contactless method. When PD activity appears on the surface of CC, it creates a typical pattern of electromagnetic field in the surrounding space of a CC. This pattern can be acquired by an antenna, as it was done in our previous works [3], [4]). The main advantage of this contactless approach is its price - antenna is much cheaper than a high voltage capacitive divider or Rogowski coil [5]. On the other hand, there are also disadvantages. The detection range is shorter and the low frequency component of a signal cannot be acquired by this method.

A. Data Acquisition Hardware Description

To obtain data from a real environment, already existing data acquisition (DAQ) platforms were used. These platforms are situated on various MV overhead power lines with CC and their original purpose is to provide PD detection via galvanic method (voltage signal evaluation) [6]. Each platform contains DAQ card with four inputs. Three inputs are being used for galvanic PD detection, the fourth input is backup. For this experiment, 4

Fig. 1.   High impedance fault on medium voltage power line.

Fig. 2.   Antenna mounted under overhead power line.

of these platforms were additionally equipped with BONI whip antennas. It is a common wide band whip antenna, used by HAM operators. The original frequency range of this antenna is 20 kHz – 300 MHz. However, because of the limited DAQ sample rate, the frequency range of the acquired signals is up to 20 MHz. The antenna outputs are connected to the backup input of PD detector DAQ, so the original functionality of the detector is preserved. This is an essential aspect, because this galvanic method is used for acquired antenna data verification.

The antenna is mounted on the pole, parallel to the CC as it can be seen in Fig. 2. Once an hour, the output signal from the
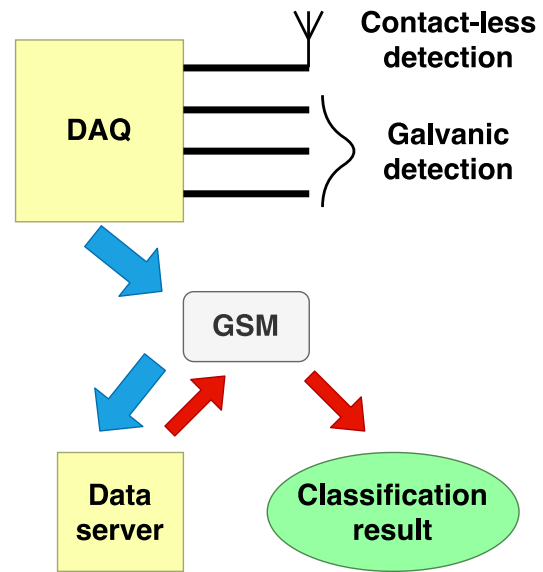

Fig. 3.   PD detection block diagram.

antenna is acquired at a sample rate of 40 MS/s in a total length of 20 ms. In a case of a tree fall, PD activity is stable and continuous, because the tree lies directly on the covered conductors. This is the most dangerous situation from the point of view of conductor insulation. Unstable PD activity is typical for random contacts with surrounding vegetation (usually because of strong winds), but these random contacts have a low impact on insulation damage. From our experience, acquiring data once per hour is sufficient for the purpose of insulation damage detection. Shorter acquisition period was also tested, but it caused trouble with slow data connection. Each acquired signal contains 800 000 samples and represents one period of the power grid frequency. Frequency range of the DAQ with connected antenna is 20 kHz – 20 MHz, the resolution is 8 bits with variable range up to 1 Vpp. Acquired signal is transmitted through GSM network to the database. The DAQ platform and the current process of PD detection are shown in Fig. 3. Because of the usage of GSM and slow data rates caused by very low signal level for distant places, it is considered to make the computations directly in DAQ instead of a data server. The proposed algorithm was developed based on this requirement.

BONI whip is an active antenna, which requires connection to the power supply for its proper function. The antenna is powered via a coaxial cable to make the installation as easy as possible (no additional wiring is needed). Nominal antenna power supply voltage is 12 V, the current is up to 150 mA. The whole acquisition platform is powered via instrument transformer (22/0.1 kV), which is connected directly to the examined power line. However, the amplifiers in DAQ platform introduces DC offset in the acquired signals. Luckily, the amplitude of this offset is very low, so it usually does not cause any significant problem.

The energy (and severity) of PD pulse is described by its apparent charge. In laboratory condition with overhead power line model connected to the PD calibrator, the sensitivity of BONI whip antenna sensor is 1 nC/10 mV. However, this value
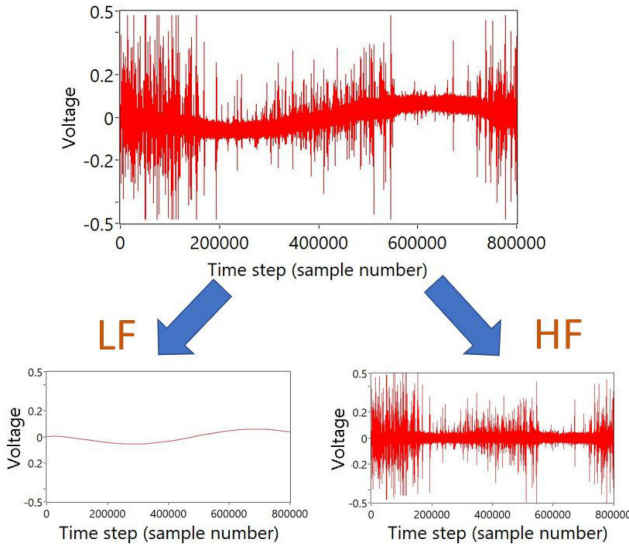
Fig. 4.   PD pattern decomposition.

depends on actual power line parameters (length, height above ground, conductor type, actual grid configuration etc.). This is why sensitivity varies in a case of real environment. The only way of sensitivity estimation in a real environment is on-site calibration. For this purpose, the whole distribution power grid would have to be shut down. This is not acceptable for distribution grids operators, so the on-site calibration was never performed.

B. PD Pattern Description

PD pattern signal from the galvanic PD detection can be decomposed into two components. Low frequency (LF) component consists of a 50 Hz carrier wave (power grid frequency), the high frequency component (HF) contains the PD activity, see Fig. 4.

Position of the HF PD peak in relation to the LF carrier wave is the key feature for PD classification [7], it is described by so-called phase resolved waveform diagrams. However, this method cannot be used for signals acquired via an antenna, because LF component is missing [5]. It is far below the lower cutoff frequency of the antenna.

The acquired signal contains only HF component without a 50 Hz carrier. This is what makes the classification of these signals so challenging.

III. CLASSIFICATION ALGORITHM

The acquired data contains a high level of background noise. As it was observed in the data, most of the background noise is within normal deviation. It is important to note that the presence of outliers does not imply the presence of the HIF. There are other types of discharges, which are not harmful to the CC insulation and need to be discerned as such. Among the other types of discharges usually present are the corona discharge, which is noticeable by the symmetric discharges in both amplitude directions and random solitary outliers.

The proposed algorithm consists of four parts: outlier detection, outlier clustering, feature extraction, and classification. The whole process is focused on making the computation as fast as possible with minimal hardware requirements. This is to enable computation on site with already installed gateway devices, which are normally used for measuring and control of the distribution power line. With fast computation time, the gateway device can conduct its original task while being able to determine the fault by themselves with minimum data transmissions. At the moment, the time series are being sent to the server for the analysis and this leads to delays in delivering the results for the areas with slow data connection. Only the outlier detection runs on the whole time series of the signal and has a computational complexity of $O(n)$. This means that the effort needed to make the computation scale linearly with the processed data size. All subsequent operations are done on a much smaller subset of the outliers, saving a lot of computational resources. Detected outliers are then clustered based on their temporal distance. Afterwards, the features are computed for individual clusters and then combined into final features common independently from the number of clusters. Finally, the classification is learnt offline on the selected features and the final model can be deployed even on embedded devices with smaller computing power. The features are based mostly on statistics and the classification itself is done by the XGBoost [8] model. The whole solution is focused on fast computation even for long time series and maximizing the precision of the algorithm, i.e., minimizing the false positives.

A. Outlier Detection

The signal measured by the wireless antenna contains a high amount of noise artefacts given mostly by AM radio and the resistance in the circuitry itself. Given that the device is always measuring the highest frequency received, the noise usually hides a large part of the original signal and creates an almost compact body when visualized. Furthermore, this compact body changes its magnitude over time. Due to this, it is not possible to define a single value which would discern outliers and noise. Another characteristic of the signal is that it is processed by the 8 b analog-to-digital converter (ADC) which returns values in the range $< -128, 128 >$.

Based on this knowledge, the outliers may be defined as all points which do not have at least c_{min} neighbours in a given area. Limits of this are given by two parameters ε and k. This area is limited by a maximal temporal distance $\varepsilon/2$ in both directions from the point. The value range for a point with value x is $[x - k, 128]$ for $X \geq 0$ and $[-128, x + k]$ for $x < 0$.

From the computational view, this can be easily implemented by counting a rolling histogram over the window of length ε and summing all values greater than $x - k$ for positive x or summing all values less than $x + k$ for negative x.

Many types of discharges which are not PD have a large number of observations in a short time span, within the whole range of values, see Fig. 5. One additional feature of this approach is that discharges such as corona are often ignored when a wide time window is used for outlier detection.
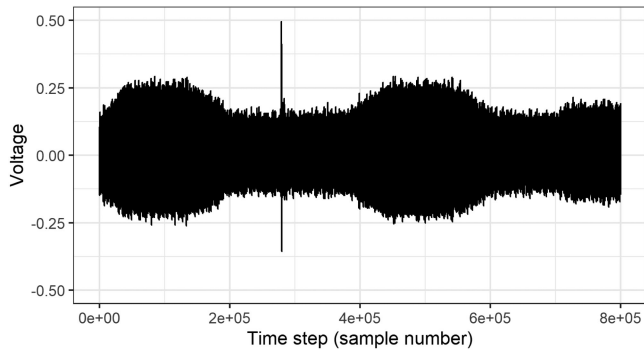
Fig. 5.   Corona discharge example.

Fig. 6.   Outlier detection and clustering example.

In our case, $c_{min} = 20$, $k = 30$, and $\varepsilon = 1001$ gave the best results and will work with the same setup as described in subsection II-A. These parameters could be modified if different hardware is used, or different kind of outliers are to be found. To make a general rule on setting these parameters is not easy since it always depends on the use case, however it is possible to provide guidance on how to approach this. The easiest rule could be said for k, which is set as a little over 11% of the maximum range of the signal, and this approach should be good in most cases. In the case when 8 b ADC is used, this range is easy to be determined based on its precision, however, even in the case of general signal a 10% of its range is a good place to start. The ε is set to 1001 based on the idea that some area around the examined point should be investigated, however not too wide, so it will catch too many other discharges. This parameter should be based mostly on the sampling frequency of the signal. In our case, the sampling frequency was 40MS/s. The last parameter c_{min} works as a sensitivity parameter. The lower value of c_{min} returns a smaller number of outliers since it is more restrictive on other spikes that may occur in the neighbourhood of the examined point. One other thing that affects this parameter is the amount of noise in the data. If the data contain less noise, it is safe to use the lower value, since there are a lower number of spikes produced by noise.

B. Outlier Clustering

After the outliers are found, a simple procedure is used to cluster the outliers into groups. In essence, the idea is to find a cluster of outliers which are close together. The clustering is done separately for the positive and negative outliers, since they are part of different discharge behaviour.

The parameters d and m are chosen, where d stands for the maximal distance of two points to be considered as part of a cluster and m is the minimal number of points grouped together to be considered a cluster. The outliers are considered a cluster if their distance is less than d or there is a path between two outliers created by such outliers. This simple method is similar to the DBSCAN clustering [9], but it is computed only in one dimension, so it can be computed in O(n) computational complexity. Example of cluster detection in a time series is shown in Fig. 6
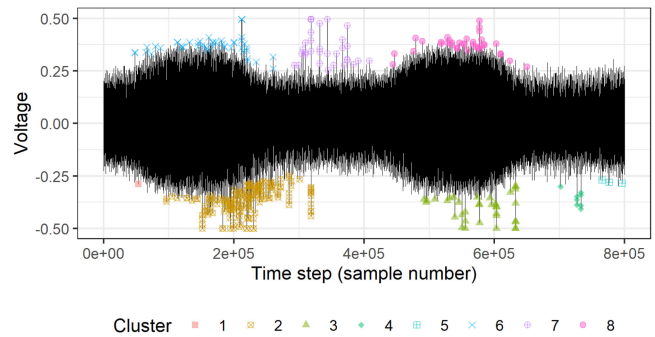
Setting of d and m should depend on the application and sampling frequency of the signal. In this case, m was set to 5, because the PDs should be defined by at least several outliers and smaller numbers would most likely mean some other kind of discharge was detected. The d parameter was set to 30 000 that equals to 3/4 of millisecond.

C. Features Computation

When the clusters are determined, the features for the classification algorithm are computed. The first stage is concerned with creating features for each cluster. Afterwards, the features are aggregated across the clusters. On the cluster level, the following 10 features are computed:

- Time range
- Standard deviation
- Minimum value
- Median value
- Mean value
- Maximum value
- Signal range
- Skewness
- Kurtosis
- Number of observations

These features were selected since based on the previous research [3]. It was found that the factors for the identification of partial discharges are the duration of PD pulse, its magnitude, and shape. However, it is not possible to apply these features in the contactless method, because of the missing phase-resolved information (LF component) in the signal. This is why the features were applied to the whole pulse-train in this experiment.

Due to the different number of clusters in each series, it is necessary to make an aggregation of these features across the clusters to create the equivalent features for every time series. The aggregation was done by computing the minimum, 1st quantile, mean, median, 3 rd quantile, and maximum of each feature across the clusters. For example, instead of 'Time range' there would be 'Minimum time range,' '1st quantile time range,' 'Mean time range,' 'Median time range,' '3 rd quantile time range,' and 'Maximum time range' variables in the aggregated feature set. These are computed from the 'Time range' variables of each cluster. This process is then executed for each of the 10 features computed for each cluster. This leads to 60 features for
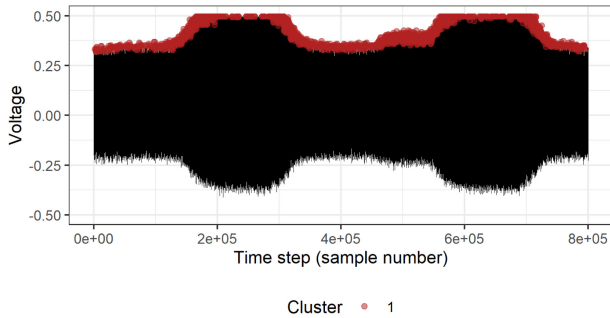
Fig. 7.   Series containing too much noise example.

the classification model. Additionally, a number of clusters and locations are added to these features. The location is a qualitative variable and therefore it needs to be transformed to quantitative variables. A dummy variable transformation was used [10], where for each unique value a new variable is created with 1 in rows where the value is true and 0 otherwise. One variable is always excluded to prevent the singularity in the feature matrix. Since the measurements were taken on 4 locations, the final model has 65 feature columns (60 + 4 locations and number of cluster variables).

D. Classification

The last stage is the classification and it is divided into three parts making preselection of the time series based on some characteristics. The first selection is whether any outlier clusters were found in the time series. If there are none, it is considered as the signal without any discharges, since there are no noticeable values pointing towards the discharge. It is not given that there will be no discharges, because they could be hidden in the noise created by external sources, however in such a case there is no way to identify the partial discharge. The second test filters the series which have too noisy measurements and it is impossible to discern between the unaffected series and the series containing partial discharge. Based on extensive analysis, it was determined that any series which contains a cluster with more than 500 observations is considered as an extremely noisy series. This characteristic was observed by many points in the positive spectra being considered an outlier in the case if there is an extreme amount of noise. It happens in the positive spectrum due to the signal offset caused by the power supply. An example of such time series with clusters is provided in Fig. 7.

If the series does not belong to neither of these categories, its features are evaluated by the model created by the XGBoost. XGBoost is a library based on the gradient boosting framework, which is known to be used in solving classification problems in industry and competitions such as Kaggle [11]. The gradient boosting is based on training multiple models and weighting the results creating an ensemble model.

E. Usage in Other Applications

The whole process is composed of several steps and most of these have some parameters. The parameters for outlier detection

and clustering are provided to increase the flexibility of the solution. For example, if one would like to detect discharges on the cables inside the factory, these would have a different pattern than the ones created by the vegetation on the overhead power lines. Therefore, it would make sense to adjust the parameters. Without precise knowledge how such discharges present themselves, it is difficult to provide a general rule for such settings. The best approach is to determine the best parameters is to use hyperparameter search. That means to search the experiment space with multiple different configurations, until a setting with good results is found. This should not be a problem in most cases since the algorithms are of linear computational complexity and therefore are computed very fast.

Another possible drawback of the proposed approach is usage of the machine learning model XGBoost, which relies on learning. This means that it is necessary to have a set of labelled data to provide for learning. In this case, we found that at least three hundred different time series had to be used for learning to provide good results. On the other hand, the advantage is that the more data are provided for learning, the better the model should be and it can improve when new data are available.

One of the reasons for making the whole process contain multiple steps is the modularity of the approach. Actually, it is possible to change, for example, the method of outlier detection for a different one and then proceed with the rest of the proposed approach. Or quite the opposite, the outlier detection may be used as a standalone part in completely different fields such as economics, musics, medicine signal analysis, etc. In the same way the clustering algorithm, feature creation, or the XGBoost model, each component may be replaced by a different one. That means the interested readers can take inspiration from our approach and replace some parts if they are unsuitable for their use case.

All functions used can be found in an R package called WIPADD.[1] This package contains individual functions for each step of the presented method as well as helper functions to make the whole process easier.

IV. RESULTS

To evaluate the proposed approach, we tested the performance on manually labelled data measured by Bony whip antenna located on overhead power lines of the Czech Republic power grid. The results of the proposed approach are compared with the current approach for the PD detection from the wireless signal presented in [4].

The Fulnecek's approach begins with data transformation. The preprocessing consists of computing the signal steepness instead of using the signal amplitude. The steepness is a simple computation of the slope of each individual increasing or decreasing part of the amplitude. Let assume P is a set of n indices of turning points of the signal amplitude series X. Then the steepness $Y(P_i)$ for $i \in [1, 2, 3, \ldots, n]$ is defined as:

$$Y(P_i) = \frac{X(P_i) - X(P_{i-1})}{P_i - P_{i-1}}. \tag{1}$$

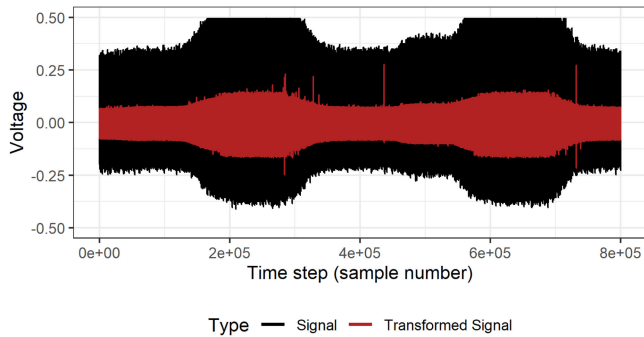[1][Online]. Available: https://code.it4i.cz/ADAS/wireless-pd-detection

Fig. 8.   Example of transformation into the steepness series.

Example of the signal amplitude and signal steepness series is shown in Fig. 8.

The Fulnecek's approach for PD detection then detected outliers by means of z-score testing. Finally, it checks whether there is a cluster of outliers which are close to each other and the size of the cluster is larger than a set parameter. In the case that such a cluster is found, the series is assumed to contain the PD. If it is not found, then the series is assumed to be without PD.

Many measures may be used to evaluate the effectiveness of the algorithm classification capabilities. These measures are usually based upon different ratios of true positive (TP), false positive (FP), true negative (TN), and false negative (FN) predictions. In the case of partial discharge detection, the most important feature of the algorithm is to reduce the false positive prediction, even at the cost of increasing the false negative predictions. This is due to the fact that sending a team to check on the cables in remote locations is costly and if the algorithm makes a lot of false positive calls, the operator will disregard the warnings given by such a system. The observed measures are:

- Precision - TP / (TP + FP)
- Sensitivity - TP / (TP + FN)
- Specificity - TN / (FP + TN)
- Accuracy - (TP + TN) / (TP + TN + FP + FN)

The aim is to maximize the sensitivity and precision. Of course, these two measures are contradicting each other, since in most cases it is possible to increase one for the cost of decreasing the other. The specificity and accuracy are less important and pretty high due to the fact that the given dataset is very imbalanced as partial discharges do not occur often. Validation of the concept was done using the R statistical software [12] and using the packages tidyverse [13] and [14] for computation and visualization.

A. Data

The dataset consists of 1700 time series of signal amplitude, with 1520 series without partial discharge and 180 series with partial discharge. All signals were checked by the human operator for the presence of partial discharges to make sure that the time series are labelled correctly. This validation was done based on data acquired by the galvanic method of PD detection. The chosen signals contained a variety of different characteristics to make the tests as general as possible. For example, the signals

TABLE I
PARAMETER SETTINGS USED FOR THE TEST

Model	Parameter	Value
Fulnecek	α	13
	d	4000
	l_{min}	10000
Proposed	c_{min}	20
	k	30
	ε	1001
	d	30000
	m	5

TABLE II
RESULTS FOR THE DATA EXCLUDED FROM THE LEARNING PART DUE TO
INSUFFICIENT KNOWLEDGE. (ABSOLUTE NUMBERS)

Model	Type	TP	FP	FN	TN
Fulnecek	Excluded	0	18	33	204
Proposed	Excluded	0	0	33	222
Fulnecek	No Groups	6	1	29	488
Proposed	No Groups	0	0	35	489

without PD contained other types of discharges, such as corona, so the model had to learn to differentiate between PDs and other discharges. Single series is 20 ms of data measured with a frequency of 40 MS/s, resulting in 800 000 observations. These series are from 4 different places in Czech Republic resulting in slightly different characteristics for each location.

B. Comparison With the State-of-the-Art

Making a fair comparison for all variants is quite difficult due to the fundamental differences of the two approaches. Fulnecek's approach is dependent only on several parameters, but does not rely on learning, while the proposed algorithm requires to learn the model before the possibility to make a classification on the new data. The learning was made on 70% of the data which had features from outlier groups and were not classified as too noisy. To make the results more informative, tests were made with 300 different random combinations of training and testing data.

In Table I there are the settings of all parameters for both methods. These were experimentally found to work the best through the hyperparameter search.

Parameter α is a cutoff for the z-score values computed in the Fulnecek's model, all the values of z-score greater than α are considered to belong to discharges, or outlier in our sense. Parameter d is the same as in our case and it determines the distance of outliers which can be considered as part of the same cluster. In Fulnecek's model, if a series contains a cluster with a temporal length of at least l_{min} it is considered that it contains a partial discharge. Parameters c_{min}, k, ε and m are described in detail in Section III. The default settings for tree model learning with logistic binary output were used for XGBoost models.

First, in Table II there is comparison for the series which were classified as too noisy ($Excluded$) and the ones which had no outlier groups ($No\ Groups$). In these cases, we will assume that

TABLE III
RESULTS FOR THE PREDICTIONS. (COUNTS)

Model	Type	TP	FP	FN	TN
Fulnecek	Best precision	16	15	20	225
Proposed	Best precision	13	0	20	243
Fulnecek	Worst precision	4	27	22	223
Proposed	Worst precision	9	13	23	231

TABLE IV
RESULTS FOR THE PREDICTIONS. (MEASURES)

Model	Type	Precision	Sensitivity	Specificity	Accuracy
Fulnecek	Best	51.61%	44.44%	93.75%	87.32%
Proposed	Best	100%	39.4%	100%	92.75%
Fulnecek	Worst	12.9%	15.38%	89.2%	82.25%
Proposed	Worst	40.91%	28.13%	94.67%	86.96%
Fulnecek	Mean	32.77%	30.98%	91.18%	83.85%
Proposed	Mean	68.9%	42.6%	97.27%	90.56%

TABLE V
RESULTS FOR THE DATA EXCLUDED FROM THE LEARNING PART DUE TO
INSUFFICIENT KNOWLEDGE. (DATA TRANSFORM, ABSOLUTE NUMBERS)

Model	Type	TP	FP	FN	TN
Fulnecek	Excluded	1	8	37	173
Proposed	Excluded	0	0	38	181
Fulnecek	No Groups	0	2	28	787
Proposed	No Groups	0	0	28	789

TABLE VI
RESULTS FOR THE PREDICTIONS. (DATA TRANSFORM, COUNTS)

Model	Type	TP	FP	FN	TN
Fulnecek	Best precision	17	14	21	147
Proposed	Best precision	28	0	15	156
Fulnecek	Worst precision	5	32	19	143
Proposed	Worst precision	18	9	6	166

there is no PD in the series as an output of the proposed model. It is clear that in the case of $Excluded$ series neither model could detect any of the 33 PD in the series, however Fulnecek's model had 18 false negative results. On the other hand, for $No\ Group$ series, the Fulnecek's approach correctly detected 6 PD out of 35 and had only 1 false positive. Overall, this process excluded 68 series with the PD either due to the PD being in the noise or being too small to notice in the original data.

The stability of the model is checked by the randomly initialized cross-validation of 300 different sets of training and validation data on the 921 series that were left. The training set contained 645 series and the validation set contained 276 series. Table III shows the counts of true positive, false positive, false negative, and true negative for the runs with best and worst precision for each model. It is clear that in the best case, the Fulnecek's model detected 16 series with PD out of 36 and has 15 false positive predictions. The proposed model could detect 13 series with PD out of 33 and had 0 false positive predictions. In the worst case, the Fulnecek's model detected 4 PDs out of 26 and 27 false positives predictions. The proposed model in the worst case detected 9 PDs out of 31 and 13 false positives predictions.

Next, the results are presented in terms of the measures shown in Table IV. In the first four rows, there is precision, sensitivity, specificity, and accuracy for the values from the Table III.

The fifth and sixth rows are the mean of the measures from the 300 samples. These are also called cross-validated measures and they give an intuition about the most likely performance in practice. The specificity and accuracy are always high, due to the high number of series without PDs. In reality, if the model predicted that there is no PD in all series, these numbers would be still quite high due to the unbalance in the dataset. Therefore, the focus is on the precision and sensitivity. With the same model, adjusting its parameters usually leads to trade-off between these two measures, since the difference is that one depends on false positives and the other on false negatives. Usually, increasing one leads to decreasing the other unless the model would get

better overall. As was already mentioned, the most important measure in this case is the precision, because it is important to have as little false positives as possible.

From the cross-validated measures, we can see that on average, the precision of the Fulnecek's model was 32.77 and the proposed model had a precision of 68.9. The proposed model also had a better sensitivity of 42.6 on average compared to 30.98. However, in case of the best precision scenarios, the Fulnecek's model had a little better sensitivity.

C. Results Using Transformed Data

Based on the results from Table II, where the Fulnecek's model could identify 6 PD while the proposed could not, another test was made. The proposed model was run on the data with the same data transformation as the Fulnecek's model. The results confirmed the theory from Fulnecek's work, that the steepness data are better to work with than the raw data from the sensor. After some testing of the parameters, it was determined that the optimal parameters for the proposed model on the transformed data are the same as on the raw data. Due to the different series being excluded and having no groups, it is not possible to make the direct comparison of the model performance with and without data transformation. In Table V there is the comparison of the proposed and Fulnecek's model on the $Excluded$ and $No\ Groups$ series as was in Table II. In this case, Fulnecek's model detected one PD out of 38 in the $Excluded$ group and had 8 false positives. $No\ Groups$ contained 28 series with PD which neither model detected. Fulnecek's model had 2 false positives out of 789 series without PD. Overall, there are 66 series with PD which are not detected by the proposed model in contrast to 68 when no data transform was made.

In this case, 1036 series are excluded and only 664 are used in the next phase. That means 465 series were used for training the models and 199 was used for validation. Table VI contains the counts of the classification for the best precision and worst precision samples as was in the Table III. The proposed model now has much better results with the best being 28 detected PDs out of 43 with no false positives. The worst results for the

TABLE VII
RESULTS FOR THE PREDICTIONS. (DATA TRANSFORM, MEASURES)

Model	Type	Precision	Sensitivity	Specificity	Accuracy
Fulnecek	Best	54.84%	44.74%	91.3%	82.41%
Proposed	Best	100%	65.17%	100%	92.46%
Fulnecek	Worst	13.51%	20.83%	81.71%	74.37%
Proposed	Worst	66.67%	75%	94.86%	92.85%
Fulnecek	Mean	33.27%	34.87%	85.47%	76.79%
Proposed	Mean	87.52%	72.7%	97.83%	93.49%

proposed model are 18 detected out of 24 and 9 false positives. The Fulnecek's model results are similar to those in Table III.

Table VII presents the same results as Table IV but for the data transformation version. The average precision for the proposed model jumped from 68.9 for no data transformation to 87.52. The sensitivity was also increased significantly from 42.6 to 72.7.

It is important to note that the final precision and sensitivity are lower due to the series which were excluded. Estimating the real precision in production is difficult since it depends greatly on the ratio of signals with high noise and in this paper only 30% of the series which get through the elimination process were used for validation. Overall, accounting for the *Excluded* and *No Group* time series in the final results would have no effect on the proposed model precision since there is no true positive or false positive observations and it would reduce the sensitivity slightly.

D. Detection Range

In this experiment, all acquired HIFs were situated up to 2 km from the detector. It is not easy to estimate the range of detection. Amplitude of HF component depends on two main factors. Humidity of the vegetation and fault-detector distance. Direct contact between CC and fresh branch with tree sap generates pulses with high repetitive rates and amplitude, which can be easily detected at a distance up to 1 km. On the other hand, contact with a dry/dead tree cannot be detected, even in a case of close distance to the sensor. It just does not generate any PD relevant pulses. There is also a strong background noise influence. Amplitude of the background noise is highly variable in the examined frequency band. This is caused by processes in the ionosphere, so it varies during the daytime. Usually, the highest level of background noise is presented during sunrise and sunset. In such situation, PD pattern is "buried" in the noise and it is difficult to detect it.

V. CONCLUSION

In this paper, a new method of particle discharge detection for the wireless sensor focused on low computational complexity and high precision was presented. One of the advantages of this approach is that it can be run directly on the gateway devices collecting data from the antenna, without the need to send data to the server. This method contains a novel approach to outlier detection, which is valid for noisy data with variable standard deviation of the noise, and a four-step process of fault detection. In the first step, the outliers of the series are detected. The second

part is the exclusion of all series which are found to be unsuitable for the classification model and are automatically labelled as not containing partial discharge. In this stage, a little over 6% of all excluded series contained PD. These PDs are usually not strong enough to be visible in the series. In the third stage, the time series is classified based upon the features computed from the outlier clusters in the series. Due to the nature of the input, the series is transformed into the steepness series rather than signal amplitudes. It was shown that this process increased the precision of classification by more than 50% in comparison to the previous state-of-the-art method. The precision is 87.52% and the sensitivity is 72.7% on the validation set. When the excluded time series are considered, the sensitivity would be lower, while the precision would remain the same. This means that the wireless detector can be used as a cheap alternative to the more expensive solutions to support a large area coverage. In the future, additional features will be considered in the pursue to further improve the quality of classification. Additionally, if the coverage increases, a multinode model could be considered to account for several time series outputs which are geographically close.

In general, the proposed contactless method is less effective when directly compared with commonly used traditional methods. For example, the galvanic PD detection method (unpublished, but deployed in production at Czech energy distributors), used for data validation in this experiment, provides better results (precision: 70%, sensitivity: 98.4%, specificity: 99.8%, accuracy: 96.7%). The effective range of the contactless detection is also shorter (4 km vs. 1 km). The main advantage of the proposed method lies in its low cost (10% of cost in comparison to galvanic detector sensors), because there are no high voltage parts. This enables to cover a large area of distribution power grid, where traditional methods of PD online monitoring are not cost-effective.

REFERENCES

[1] P. Pakonen, "Detection of incipient tree faults on high voltage covered conductor lines," Ph.D. dissertation, Tampere Univ. Technol., Tampere, Finland, 2007. [Online]. Available: https://trepo.tuni.fi/bitstream/handle/10024/114541/pakonen.pdf
[2] 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 Trans. Dielect. Elect. Insul.*, vol. 27, no. 6, pp. 1861–1888, Dec. 2020.
[3] S. Misak, J. Fulnecek, T. Jezowicz, T. Vantuch, and T. Burianek, "Usage of antenna for detection of tree falls on overhead lines with covered conductors," *Adv. Elect. Electron. Eng.*, vol. 15, no. 1, 2017.
[4] J. Fulnecek and S. Misak, "A simple method for tree fall detection on medium voltage overhead lines with covered conductors," *IEEE Trans. Power Del.*, vol. 36, no. 3, pp. 1411–1417, Jun. 2021.
[5] J. M. Martínez-Tarifa, J. A. Ardila-Rey, A. Prieto-Duque, J. M. Fresno, and G. Robles, "Statistical analysis of rf signals emitted by partial discharges," in *Proc. IEEE Int. Conf. Dielect.*, 2016, pp. 536–539.
[6] S. Mišák and V. Pokorný, "Testing of a covered conductor's fault detectors," *IEEE Trans. Power Del.*, vol. 30, no. 3, pp. 1096–1103, Jun. 2015.
[7] N. D. Jacob and W. M. McDermid, "Experience with partial discharge measurements on instrument transformers in high voltage laboratory acceptance tests," in *PES T. D.* 2012, pp. 1–4.
[8] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2016, pp. 785–794.
[9] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discov. Data Mining*, 1996, pp. 226–231.

[10] M. Kuhn and K. Johnson, *Feature Engineering and Selection*. New York, NY, USA: Chapman and Hall/CRC, 2020.

[11] "Kaggle," Tampere, Finland, 2020. [Online]. Available: https://www.kaggle.com/

[12] R Core Team, *R: A Language and Environment for Statistical Computing, R. Foundation for Statistical Computing*, Vienna, Austria, 2020. [Online]. Available: https://www.R-project.org/

[13] H. Wickham *et al.*, "Welcome to the tidyverse," *J. Open Source Softw.*, vol. 4, no. 43, 2019, Art. no. 1686. [Online]. Avilable: https://joss.theoj.org/papers/10.21105/joss.01686#

[14] T. Chen *et al.*, *xgboost: Extreme Gradient Boosting*, 2020. [Online]. Available: https://CRAN.R-project.org/package=xgboost

Jan Fulneček was born in Karviná, Czech Republic, in 1989. He received the Ph.D. degree from the Department of Electrical Power Engineering, Technical University of Ostrava, Moravskoslezsky, Czech Republic, in 2018. He is currently a Junior Researcher with the ENET Research Centre and as an Assistant Professor with the Department of Electrical Power Engineering, VŠB Technical University of Ostrava. His research interests include on electrical equipment diagnostics.

Tomáš Martinovič received the Ph.D. degree in computer science from the VSB – Technical University of Ostrava, Moravskoslezsky, Czech Republic, in 2018. He is currently a Researcher with team of Advanced Data Analysis and Simulation Lab, IT4Innovations, VŠB-Technical University of Ostrava. His research interests include operational research, time series analysis, machine learning, deep learning, and non-linear analysis of dynamical systems.