

Article

Classification of Superimposed Partial Discharge Patterns

Benjamin Adam * and Stefan Tenbohlen * 

Institute of Power Transmission and High Voltage Technology, University of Stuttgart, 70569 Stuttgart, Germany

* Correspondence: benjamin.f.adam@googlemail.com (B.A.); stefan.tenbohlen@ieh.uni-stuttgart.de (S.T.)

Abstract: Phase resolved partial discharge patterns (PRPD) are routinely used to assess the condition of power transformers. In the past, classification systems have been developed in order to automate the fault identification task. Most of those systems work with the assumption that only one source is active. In reality, however, multiple PD sources can be active at the same time. Hence, PRPD patterns can overlap and cannot be separated easily, e.g., by visual inspection. Multiple PD sources in a single PRPD represent a multi-label classification problem. We present a system based on long short-term memory (LSTM) neural networks to resolve this task. The system is generally able to classify multiple overlapping PRPD by while only being trained by single class PD sources. The system achieves a single class accuracy of 99% and a mean multi-label accuracy of 43% for an imbalanced dataset. This method can be used with overlapping PRPD patterns to identify the main PD source and, depending on the data, also classify the second source. The method works with conventional electrical measuring devices. Within a detailed discussion of the presented approach, both its benefits but also its problems regarding different repetition rates of different PD sources are being evaluated.

Keywords: partial discharge; PD; classification; neural networks; LSTM



Citation: Adam, B.; Tenbohlen, S. Classification of Superimposed Partial Discharge Patterns. *Energies* **2021**, *14*, 2144. <https://doi.org/10.3390/en14082144>

Academic Editor: Pawel Rozga

Received: 2 March 2021

Accepted: 6 April 2021

Published: 12 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Partial discharge measurement or monitoring is an important tool for the assessment of electrical insulation in high voltage equipment. In order to examine the PD, data phase resolved partial discharge patterns (PRPDs) are created from the measurement data [1]. An expert can determine the type of defect that has caused this particular pattern by visual inspection due to personal experience. In order to facilitate continuous condition monitoring, this workflow must be automated. Computed aided classification of partial discharges is a well-studied field [2–5]. However, most of these publications only focus on single class classification. Thus, only a single active source can be classified correctly. Classification of multiple simultaneously occurring PD has not been solved until today. The countless possible combinations of overlapping PD patterns make it difficult to train a system adequately.

This work presents a novel approach using exclusively single PD sources for the training of the classification system. This system, however, is able to classify two overlapping PD sources. Instead of using the PRPD and extracting visual features for use in the classification task, we use the sequence of PD events, which is the basis for constructing the PRPD. This sequence contains additional timing data of each PD event, so the different sources can be separated by time. Sequence classification is a widely studied topic in machine learning. In recent years, many improvements have been made for the use in voice recognition tasks. In this work, similar methods for solving the problem of PD source classification are applied. We proposed a similar approach before in Adam et al. [6]. The waveform of single PD signals was used as input for a sequence classification in this previous work.

The differences of the previous and this new method will be explained in Section 2. Section 2 will also present the used PD data of artificial laboratory sources. Section 3 illustrates the architecture of the LSTM neural network. In Section 4, the results of the

multi-label classification are presented. The paper concludes with a discussion in Section 5 and a summary of the results in Section 6.

2. PD Dataset

PD data are essential for the training and test of the classifier. In this research, artificial PD sources were recorded in the laboratory using an IEC 60270 [7] compliant measurement setup. The usage of defined artificial sources offers the advantage that the class labels (e.g., *CoronaHV*, Figure 1) are known from the beginning. Data provided by actual high voltage equipment are not suitable to be used for supervised learning because the defect types are not known in most cases. In particular, it is difficult to obtain the required amount of labeled training data from non-laboratory sources.

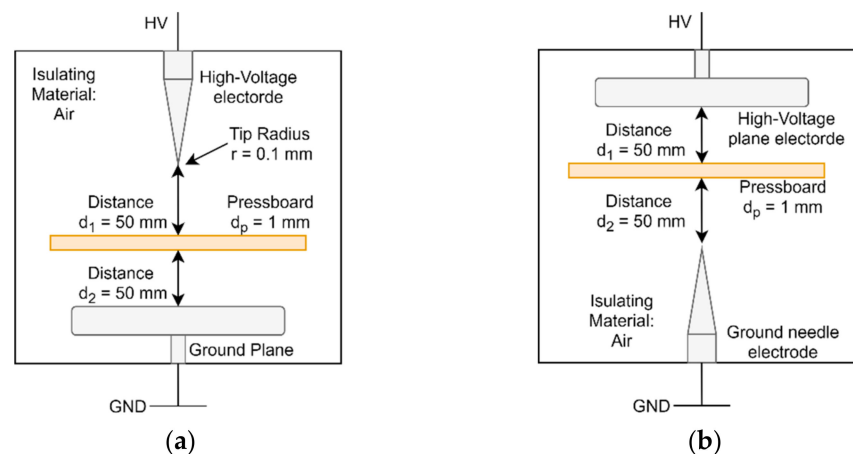


Figure 1. Construction of artificial PD sources: (a) *CoronaHV*; (b) *CoronaGND*.

2.1. Measurement Setup

Six artificial PD sources were measured in the laboratory to create the training and test data. These sources were two discharge sources in air and four sources in mineral oil (see Table 1).

Table 1. Artificial PD sources and configurations.

PD Source (Class Label)	Voltage/kV (RMS)	Time/min	N Impulses	Repetition Rate/N per s
<i>CoronaHV</i>	5.4	65	61,200	15.7
<i>CoronaGND</i>	5.1	66	265,200	67.0
<i>SP</i>	24.9	65	1,617,300	414.7
<i>GL</i>	2.9	63	2,100,700	555.7
<i>UP</i>	21.9	67	229,000	57.0
<i>SP-Stahl</i>	4.0	61	1,158,748	316.6

The artificial PD sources in air are named *CoronaHV* and *CoronaGND*. Both use a needle-plane configuration. Regarding *CoronaHV*, the needle is on high voltage potential and the plane is on ground potential. Referring to *CoronaGND*, the inverse situation is given: the plane is set on high voltage potential and the needle on ground potential. This mirrored setup generates two very distinct PRPD patterns. The dimensions, distances, and tip radius are provided in Figure 1.

The sources *SP*, *GL*, *UP*, and *SP-Stahl* are artificial PD sources measured while submerged in mineral oil. *SP* is a needle-sphere setup. *GL* describes a surface discharge where a sharp edge generates PD tangential to a pressboard plate. *UP* was constructed as a floating potential with a metal part, neither connected to high voltage nor to ground. *SP-Stahl* describes a different needle-sphere configuration in another oil-filled enclosure.

This configuration uses steel electrodes and a steel enclosing for the oil tank in comparison to the brass electrodes in the acrylic glass tank used in SP, GL, and UP. Dimensions of SP and SP-Stahl are the same as shown in Figure 2a. All sources including their respective dimensions, distances, and angles are shown in Figure 2.

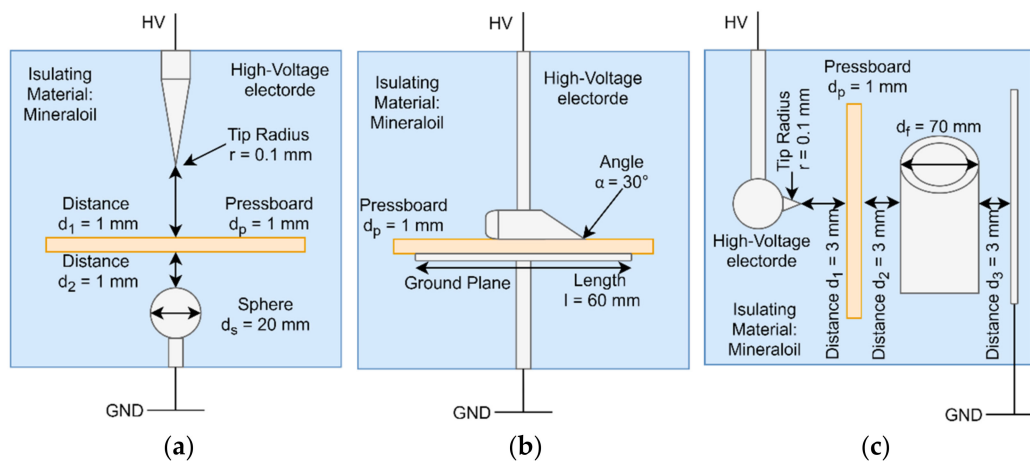


Figure 2. Construction of artificial PD sources: (a) SP & SP-Stahl; (b) GL; (c) UP.

All measurements were performed according to IEC 60270 [7]. PD impulses were integrated between $f_1 = 100$ kHz and $f_2 = 400$ kHz to get the apparent charge values. All sources were measured at their inception voltage for a duration of over 60 min (see Table 1). Complete data streams including time, apparent charge, and phase angle of each discharge were recorded. These provide the basis for the construction of a PRPD.

For the sources that used oil as insulating material, Nynas Nytro Lyra X mineral oil (Stockholm, Sweden) was used. No synthetic esters were used for the measurements.

2.2. Definition of Used Terms

Subsequently, different terms for the measurement data in its various forms are used. These terms will be defined in this section. Each actual discharge event is recorded by the measurement device by three attributes:

1. The time t when the discharge occurred,
2. The apparent discharge Q in pC, and
3. The phase angle φ .

This triplet of data defines a PD event. Each event originates from a single PD source and can have exactly one class label (i.e., only one true type of source it originated from).

Multiple events that follow each other build a sequence. This sequence is the basis for constructing a PRPD. The measurement device saves these sequences for each measurement. Each sequence is divided into smaller subsequences of 100 PD events each. Such a sequence of 100 events is called a sample (see Figure 3b). This is a sample in the machine learning sense because it is the smallest piece of data that is used as input for the classification algorithm. The 100 event sample size applies for both training and later on in the testing stages. As a sample consists of multiple PD events and (considering general monitoring data) each event can originate from a different PD source, a sample can have multiple class labels. Only for the training data were samples with a single class label used. For the testing samples with PD, events from multiple sources were used (see Figure 3c). Sections 2.3 and 2.4 describe in detail the training and the test data and their creation.

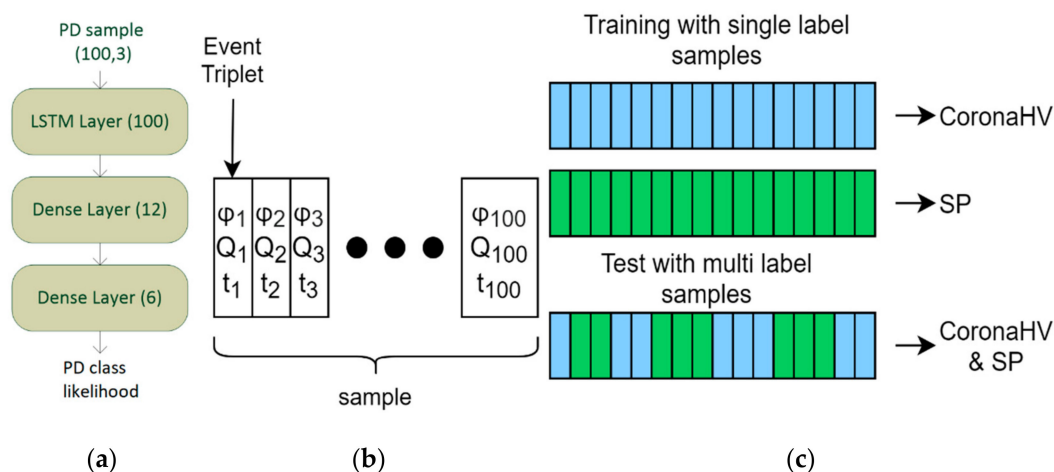


Figure 3. Neural network and input data: (a) Neural network architecture, (b) sample construction, and (c) training and test.

Side note: The definition of a sample used in this context is not to be confused with a measurement sample from a digital oscilloscope. These two completely different concepts just share the same name.

The sequence of PD events as input data stands in contrast to the signal waveforms used in Adam et al. [6]. While a waveform represents one single PD event, this event sequence uses multiple events with their respective timings. This information is not available if single waveforms are used. In addition to this, most commercial PD measurement software save the sequence of events but not the single waveforms. Therefore, this new method can be used more easily with available datasets.

2.3. Training Data

Most classification approaches use the PRPD or some features derived from the PRPD as input data for the classification. As shown in Adam et al. [8], the PRPD was originally intended as an interface for human interpretation and actually reduces the amount of available information because it accumulates all discharges over a given period of time and consequently loses all timing information. In this work, we use the data streams previously tested in [8] as training data. A data stream consists of multiple entries n of a triplet of information. This triplet contains t , Q , and φ (see Section 2.2).

For the training data, data streams of length $l_d = 100$ were extracted from the measurement data of a single source. Generally, all artificial sources differ significantly in their number of discharges, meaning they provide different repetition rates (see Table 1). In order to get a balanced training dataset, up to 2000 samples per class were extracted, which results in 200,000 single events as the maximum. This limit is introduced by source *CoronaHV* (see Figure 1), since this is the source with the lowest repetition rate (see Table 1). Only 50% of the recorded measurement data was used for the construction of training samples. The remaining data are used separately for the testing stage. This is done to prevent the neural network overfitting on the training data.

2.4. Artificial Test Data

Superimposed patterns were created for the test data. Half of the measurement samples were used in this process. First, all samples of a single source PD data were normalized on the time axis, so each sample started at time 0. Then, two samples from different sources were combined and sorted by the time axis. Of the resulting sample, only the first 100 PD events were used as a new sample. This new sample now had PD events of two different sources and, therefore, the whole sample had two class labels. With this we created $\frac{n!}{(n-k)!} = \frac{6!}{(6-2)!} = 30$ different combinations of samples with two class labels. Since the order of the labels is not important (e.g., SP-GL and GL-SP are the same), this

was reduced to 15 different combinations. For each of these combinations, between 1000 and 2000 samples were created.

A problem with this approach is the different repetition rate of different PD sources. While some sources are very active and produce many discharges in very short amounts of time, other sources only produce a few discharges in minutes. This leads to an imbalance in the test dataset, which impacts the results as shown in Section 4.3.1. This problem could be remedied by creating a balanced dataset for the test data. However, as real PD defects in high voltage equipment also exhibit different sources with different repetition rates, using imbalanced data is a better representation of actual real-life problems. We chose the imbalanced repetition rate data to test the worst-case scenario for this approach.

3. LSTM Classification

In this chapter, the classifier is introduced. The basis of the used LSTM networks is explained, and the used architecture is presented. Training and test of the classification process are described in Sections 3.4 and 3.5, respectively.

3.1. Multi-Label Classification

Multiple superimposed PD sources are a type of multi-label classification problem. This is not to be confused with multi-class classification, where the output class can be one of many classes. In multi-label classification, the output can consist of many labels at the same time [9].

In general, classification returns a single output class. If three different classes A, B, C are possible, then the output might be one of this three: [1,0,0]; [0,1,0]; or [0,0,1]. In multi-label classification, the output of three classes A, B, C can be one of the following: [1,0,0]; [0,1,0]; [0,0,1]; [1,1,0]; [0,1,1]; [1,0,1]; or [1,1,1].

There are two main ways to achieve this multi-label output. The first method uses training data for all possible combinations of superimposed classes. Therefore, the training data already contain all scenarios, and each multi-label class can be translated into a new single-label class, which represents both labels. This approach becomes difficult for high numbers of classes. For M classes, 2^{M-1} possible cases need to be considered and added to the training dataset. This also disregards all the possible fluctuations that can be found in PD measurement data.

The second method uses a modified classifier to output multiple classes if found. This can be achieved by outputting the class probabilities and choosing multiple labels from them [10]. We chose this approach in this research by modifying the last layer in the neural network. This modification is presented in Section 3.3.

3.2. LSTM Basics

Neural networks are often used in the classification of partial discharges [11,12]. Neural networks have been successfully used to solve a wide range of problems. In the last few years, big advances have been made in the research of neural networks. Most of these advances have not been introduced in PD research, yet. Often a three layer fully connected neural network is utilized in PD analysis [13]. This kind of network uses feature vectors extracted from the PD data as inputs and returns one of the possible PD source classes as output. Other classification schemes, which are not based on neural networks, work similarly in using feature vectors as inputs and calculating an output class. However, in contrast to other classifiers like random forest or support vector machines, three-layer neural networks have significant disadvantages. They require a lot of data for effective training and training is time consuming [14]. When working with feature vectors, alternative classification algorithms should be considered.

Neural networks, especially modern deep neural networks, consist of very particular architectures that are used to solve a small set of specific problems. They cannot be used as a general tool to solve every kind of problem. Many of these problems lay in the domain of image classification. They can be solved by using convolutional neural networks. In this

work, we want to solve a sequence classification problem. This kind of problem can be solved by recurrent neural networks and by extension LSTM networks.

The classifier is constructed on the base of long short-term memory (LSTM) neural networks. LSTM are a type of recurrent neural networks (RNN). RNN are networks in which the neurons loop back on themselves. They are used to classify sequence data. RNN have the problem that for long sequences, the training fails. This is called the vanishing gradient problem [15]. LSTM solves this problem and can be used to classify long data sequences. Additional information on LSTM networks, their internal mechanisms, and cell structure can be found in [15]. In this work, we use LSTM networks as sequence classifiers.

3.3. Neural Network Architecture

A three-layer architecture was chosen (see Figure 3a). The first layer is an LSTM layer with 100 neurons. The second layer is a dense layer with twelve neurons and a sigmoid activation function. The output layer consists of another dense layer with six neurons, which correspond to the six output classes. Normally a softmax activation function is used for the last layer in order to obtain the final output class label. The softmax function sets the highest neuron value in the layer to 1 and all other values to 0. This is useful if only the class with the highest confidence should be returned. This, however, makes it impossible to classify multiple superimposed PD classes since only the class with the highest output value is selected as class label.

Instead, we use a sigmoid activation function to return the class probability for each of the six classes. A threshold is then used to assign class labels. If the class probability exceeds the threshold, the class is recognized as present in the data. In this manner, multiple classes can be found in one sample of PD data. A scheme to verify this multi-label is presented in Section 4.1. Different threshold values were tested in order to find a good value for our dataset. All tested thresholds and their classification accuracy are presented in Section 4.2.

The neural network was implemented in Python 3.6 with Keras [16] and the TensorFlow backend [17]. For the loss function binary crossentropy was used, and for the optimizer adam was used. The network was trained for 100 epochs with a batch size of 50.

3.4. Training

Six classes were used for the training of the classifier. The training dataset from the measurement (see Section 2) was used. For the training, only the single classes were used. No data from superimposed PD sources were used in the training process. After the split of the dataset into 50% training data and 50% test data, only 90% of the training data were used for the training. This means 78,693 data points were used. The remaining 10% of the training data (8744 data points) were used for a verification test. This showed the classifier worked with an accuracy of 99% for single class samples.

3.5. Test

After the single class verification, the test dataset created in Section 2.4 was used. The test dataset consists of superimposed PD sources of all possible combinations. The data used for creating the test dataset were not previously used for the training dataset. The results of the test with different thresholds are presented in Section 4.

4. Results and Validation

This section describes the results of the multi-label classification of superimposed PD data. First, the classification accuracy for multi-label problems is defined. Then the different threshold values are compared in regards to their influence on the total classification accuracy. In Section 4.3, the classification results for all combinations of superimposed patterns are discussed. The classification does not work equally well for every combination. The cause for this behavior is discussed.

4.1. Classification Accuracy for Multi-Label Problems

In order to evaluate and compare different classification schemes, different assessment criteria are used. One widely used metric is the classification accuracy. In single-label predictions, the accuracy represents the ratio of correct prediction over the total number of predictions.

Single-label classification accuracy:

$$\text{accuracy single-label, } A_{sl} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (1)$$

Equation (1) shows the single-label accuracy; TP stands for True Positives, TN for True Negatives, FP for False Positives, and FN for False Negatives in the confusion matrix. Regarding multi-label classification, this definition does not suffice anymore. The definition of True Positive is not easy to apply since a multi-label can be partially correct, partially wrong, or incomplete. One possibility is using the Exact Match Ratio where partially correct predictions are considered false and only exact matches are considered true. We chose the accuracy definition by Godbole & Sarawagi [18] for multi-label problems (see Equation (2)). Here, accuracy is defined as the proportion of the correctly predicted labels to the total number of labels for each instance. To get the accuracy for all samples, the mean of all predicted instances is taken.

Multi-label classification accuracy:

$$\text{accuracy multi-label, } A_{ml} = \frac{1}{n} \sum \frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|}, \quad (2)$$

where Y_i are the correctly predicted labels, Z_i are the ground truth labels, and n is the number of predictions. An accuracy value of 1.0, for example, would mean a perfect match of predicted labels and ground truth labels, whereas an accuracy value of 0.5 represents that half of the multi-label predictions were correct.

4.2. Threshold

For the last layer of the neural network, different threshold values were chosen. Traditional wisdom would suggest a class probability of 0.5 or greater to be a good criterion for assigning a label to a sample. In reality, a much smaller threshold showed much better classification accuracies (see Table 2).

Table 2. Classification accuracy for different threshold values.

Threshold	Mean Classification Accuracy
0.50	0.37
0.10	0.42
0.05	0.43
0.01	0.40

Normally, the last layer in the neural network outputs a class probability for each possible class. The softmax function chooses the neuron with the highest numerical value as output for the class prediction. This was not used here because softmax returns only one class and we need the possibility of multiple classes. Instead, the sigmoid function was used to return a class probability for each of the six classes as explained in Section 3.3. These class probabilities are backpropagated in the training phase in order to update the weights in the neurons of the network. For this process, the class probability in the last layer does not have to surpass a specific value. It just must be higher than all the other classes. If all other class probabilities are close to zero, even small values in the neurons for the true classes are sufficient for successful backpropagation.

Figure 4 shows an example of class probabilities for two superimposed PD sources. It can be seen that one class has a very high probability and the second class shows a much

lower probability. All other classes are close to zero. In order to detect multiple classes, a threshold must be chosen, which can detect the second, smaller class probability. After preliminary inspection of the data in form of figures like Figure 4, it could be seen that most probability values for the second most probable class were between 0.05 and 0.1. Therefore, these values were chosen for accuracy testing. Additionally, 0.5 was chosen as the naive assumption and 0.01 was chosen as an extreme value in the other direction.

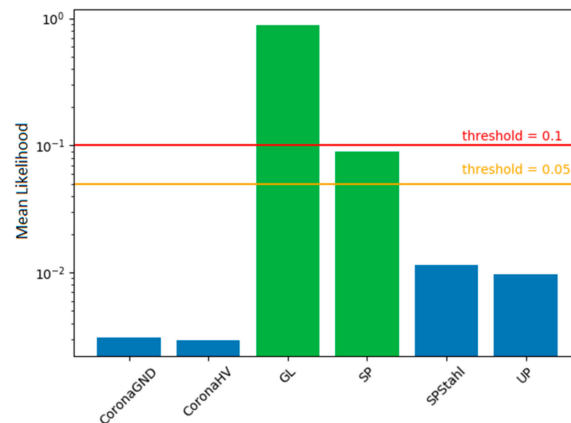


Figure 4. Class probabilities of two superimposed PD sources (log scale).

As seen in Table 2, a threshold of 0.05 showed the best overall classification accuracy for multi-label data. Because of computational time limitations and the need to retrain the complete network for every threshold, only these four values were tested.

4.3. Classification of Superimposed PD Data

With the chosen threshold of 0.05, overall classification accuracy of 43% was achieved for multi-label data (see Table 2). That means that almost half of the samples could be classified correctly with two labels. Table 3 shows that not all combinations show the same high accuracy. Combinations *CoronaGND-CoronaHV*, *GL-SP-Stahl*, and *UP-CoronaHV* show lower classification accuracy than all the other variants, whereas combinations *GL-UP*, *GL-SP*, *SP-Stahl-SP*, and *SP-Stahl-UP* show the highest accuracy for multi label classification. Table 3 shows the classification accuracy and the average PD events per class for all samples of the selected label combination.

Table 3. Multi-class accuracy and mean number of events.

Label Combination		Mean Number of Events of Class 1	Mean Number of Events of Class 2	Accuracy Score
Class 1	Class 2			
<i>CoronaGND</i>	<i>CoronaHV</i>	87.54	14.23	0.00
<i>CoronaGND</i>	<i>UP</i>	61.21	39.59	0.24
<i>GL</i>	<i>SP-Stahl</i>	97.42	2.58	0.26
<i>UP</i>	<i>CoronaHV</i>	23.20	77.97	0.29
<i>CoronaGND</i>	<i>SP-Stahl</i>	44.02	58.14	0.34
<i>CoronaHV</i>	<i>SP-Stahl</i>	13.73	88.03	0.38
<i>CoronaGND</i>	<i>SP</i>	17.24	89.62	0.47
<i>CoronaHV</i>	<i>SP</i>	7.13	97.38	0.47
<i>SP</i>	<i>UP</i>	90.62	10.16	0.48
<i>CoronaGND</i>	<i>GL</i>	4.16	98.92	0.49
<i>CoronaHV</i>	<i>GL</i>	7.86	99.71	0.49
<i>GL</i>	<i>UP</i>	98.50	2.50	0.52
<i>GL</i>	<i>SP</i>	90.21	9.79	0.56
<i>SP-Stahl</i>	<i>SP</i>	79.55	20.50	0.67
<i>SP-Stahl</i>	<i>UP</i>	64.96	35.13	0.72

4.3.1. On the Topic of Imbalanced Data

Different factors might be responsible for the different classification accuracies of the label combinations. Since all classes in the data show different repetition rates of PD events, this could have been a significant factor. If a sample consists almost entirely of only PD events of one class, the classification of the other class will not be possible. Table 3 shows that some of the label combinations with small accuracy values have highly imbalanced repetition rates (e.g., *GL-SP-Stahl*), while other label combinations like *GL-SP* show higher accuracy scores. This leads us to the conclusion that the imbalance of data and repetition rates are not the main mechanism causing the accuracy differences.

5. Discussion

The mean accuracy of 43% seems rather low compared to previous works like [6] where an accuracy of 97% was achieved on the same dataset. As explained in Section 4.1, a different accuracy score for multi-label classification was used. There are multiple possible explanations for different values of this score. A score of 0.5 for example could mean that one class was always classified correctly while the second class was always classified wrongly. However, it could also mean that both classes were classified correctly half of the time. Our score of 43% could be seen as an equivalent single label score of 86%, which is closer to our previously achieved accuracy for the single-label problem. Unfortunately, the multi-label accuracy metric does not provide a detailed comparison between single- and multi-label performance.

This is also very dependent on the input data and class combination itself, as seen in Table 3. If we take a look at the combination of *CoronaHV* and *GL*, we see that most of the data are from the *GL* class. This is a direct result of the much higher repetition rate of *GL* as shown in Table 1. The multi-label accuracy from this specific class combination is 0.49. We interpret this as almost all instances of *GL* are classified correctly (0.5 would be a perfect score for the first class), while none of the instances of *CoronaHV* are detected. In contrast, class combination with similar repetition rates (for example *CoronaGND* and *UP*) achieve lower multi-label accuracy scores.

This suggests that for single-label data or instances where one class provides most of the data for the sample, we get comparable, albeit slightly lower, accuracy scores. In some cases, however, the second class can be additionally classified correctly.

While the threshold of 0.05 provides good results, further investigation is needed to find better threshold values. A grid search between 0 and 1 with a step size of 0.01 could be used to iterate over a bigger value range. Since this requires retraining of the neural network and computational time was limited for this work, this needs to be done in further studies.

Different combinations of sources result in different multi-label accuracies. While it seems that this result is not dependent on the differences in repetition rate, this relationship needs to be investigated further.

6. Conclusions

We showed that superimposed PRPD pattern, with two PD sources present, can be classified by use of an LSTM neural network with a mean multi-label accuracy of 43%. Temporal information is used in the form of a sequence of PD events, measured according to IEC 60,270, instead of the image form of PRPD patterns. Sequence classification is done by LSTM networks. The neural network is modified to support multi-label output. With a threshold level on the network's last layer, a sequence can be classified as containing multiple sources at the same time. For the training of the neural network, only single class PD data are used. For single source PD, this method achieves a classification accuracy of 99%. For two superimposed PD sources, the method achieves a mean classification accuracy of 43% for a threshold of 0.05 and the worst-case scenario of an imbalanced dataset. Depending on the specific PD types, the classification accuracy fluctuates widely. This problem could be improved by using longer sequences as input data. In the future,

this method can also be used for PD classification in other equipment. For DC applications, a variation without the phase angle information could also be possible. In this version, only charge and time information would be used.

Author Contributions: Conceptualization, B.A.; investigation, B.A.; data curation, B.A.; writing—original draft preparation, B.A.; writing—review and editing, B.A., S.T.; visualization B.A.; supervision, S.T. Both authors have read and agreed to the published version of the manuscript.

Funding: We thank the “Deutsche Forschungsgemeinschaft” (DFG) for funding this project (Projekt-nummer: 317310948).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. CIGRE WG D1.11. *Knowledge Rules for Partial Discharge Diagnosis in Service*; Technical Brochure 226; International Council on Large Electric Systems: Paris, France, 2003.
2. Sahoo, N.C.; Salama, M.M.A.; Bartnikas, R. Trends in partial discharge pattern classification: A. survey. *IEEE Trans. Dielectr. Dielectr. Electr. Insul.* **2005**, *12*, 248–264. [[CrossRef](#)]
3. Wu, M.; Cao, H.; Cao, J.; Nguyen, H.L.; Gomes, J.B.; Krishnaswamy, S.P. An Overview of State-of-the-Art Partial Discharge Analysis Techniques for Condition Monitoring. *IEEE Electr. Insul. Mag.* **2015**, *31*, 22–35. [[CrossRef](#)]
4. Raymond, W.J.K.; Illias, H.A.; Bakar, A.H.A.; Mokhlis, H. Partial Discharge Classifications: Review of Recent Progress. *Measurement* **2015**, *68*, 164–181. [[CrossRef](#)]
5. Engel, K.; Kurrat, M.; Peier, D.; Schnettler, A. Klassifikation von inneren Teilentladungen mit neuronalen Netzen. *Elektrizitätswirtschaft* **1993**, *93*, 319–323.
6. Adam, B.; Tenbohlen, S. Classification of multiple PD Sources by Signal Features and LSTM Networks. In Proceedings of the 2018 IEEE International Conference on High Voltage Engineering and Application (ICHVE), Athens, Greece, 10–13 September 2018.
7. IEC. *IEC 60270 High-Voltage Test Techniques—Partial Discharge Measurements*; International Electrotechnical Commission: Geneva, Switzerland, 2000.
8. Adam, B.; Tenbohlen, S. Vergleich von Merkmalen und Klassifikatoren für eine automatisierte Auswertung von Teilentladungsmustern. In Proceedings of the VDE-Fachtagung Hochspannungstechnik 2016, Berlin, Germany, 14–16 November 2016.
9. Tsoumakas, G.; Katakis, I.; Vlahavas, I. Mining Multi-Label Data. In *Data Mining and Knowledge Discovery Handbook*, 2nd ed.; Maimon, O., Rokach, L., Eds.; Springer: Berlin/Heidelberg, Germany, 2010.
10. Spolaôr, N.; Cherman, E.A.; Monard, M.C.; Lee, H.D. Comparison of Multi-label Feature Selection Methods using the Problem Transformation Approach. *Electron. Notes Theor. Comput. Sci.* **2013**, *292*, 135–151. [[CrossRef](#)]
11. Mehrotra, K.; Mohan, C.K.; Ranka, S. *Elements of Artificial Neural Networks*; MIT Press: Cambridge, MA, USA, 1997; p. 344.
12. Gulski, E. Computer-aided measurement of partial discharges in HV equipment. *IEEE Trans. Electr. Insul.* **1993**, *28*, 969–983. [[CrossRef](#)]
13. Schmidhuber, J. Deep Learning in Neural Networks: An Overview. *Neural Netw.* **2015**, *61*, 85–117. [[CrossRef](#)]
14. Mijwel, M. Artificial Neural Networks Advantages and Disadvantages. 2018. Available online: https://www.researchgate.net/publication/323665827_Artificial_Neural_Networks_Advantages_and_Disadvantages/citation/download (accessed on 12 March 2021).
15. Hochreiter, S.; Schmidhuber, J. Long Short-term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
16. Chollet, F. “Keras” 2015. Available online: <https://keras.io> (accessed on 12 March 2021).
17. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Software. *arXiv* **2016**, arXiv:1603.04467. Available online: <https://tensorflow.org> (accessed on 12 March 2021).
18. Godbole, S.; Sarawagi, S. Discriminative Methods for Multi-labeled Classification. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 22–30.