

Article

Online Predictive Maintenance Monitoring Adopting Convolutional Neural Networks

Christian Gianoglio , Edoardo Ragusa , Paolo Gastaldo , Federico Gallesi  and Francesco Guastavino 

Electrical, Electronics and Telecommunication Engineering and Naval Architecture Department (DITEN), University of Genoa, 16145 Genova, Italy; edoardo.ragusa@edu.unige.it (E.R.); paolo.gastaldo@unige.it (P.G.); federico.gallesi@edu.unige.it (F.G.); francesco.guastavino@unige.it (F.G.)

* Correspondence: christian.gianoglio@edu.unige.it

Abstract: Thermal, electrical and mechanical stresses age the electrical insulation systems of high voltage (HV) apparatuses until the breakdown. The monitoring of the partial discharges (PDs) effectively assesses the insulation condition. PDs are both the symptoms and the causes of insulation aging and—in the long term—can lead to a breakdown, with a burdensome economic loss. This paper proposes the convolutional neural networks (CNNs) to investigate and analyze the aging process of enameled wires, thus predicting the life status of the insulation systems. The CNNs training does not require any kind of assumption of how the factors (e.g., voltage, frequency and temperature) contribute to the life model. The experiments confirm that the proposal obtains better estimations of the life status of twisted pair specimens concerning existing solutions, which are based on strong hypotheses about the life model dependency on the factors.

Keywords: predictive maintenance; convolutional neural networks; partial discharges



Citation: Gianoglio, C.; Ragusa, E.; Gastaldo, P.; Gallesi, F.; Guastavino, F. Online Predictive Maintenance Monitoring Adopting Convolutional Neural Networks. *Energies* **2021**, *14*, 4711. <https://doi.org/10.3390/en14154711>

Academic Editor: Wojciech Sikorski

Received: 17 June 2021
Accepted: 29 July 2021
Published: 3 August 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

The online monitoring of high voltage (HV) apparatuses prevents economic losses due to the breakdown of an insulation system [1,2]. Usually, the conditions of an electrical machine are assessed via periodical checks, which cause a temporary disservice and a waste of money. Conversely, online predictive maintenance estimates the status of the insulation system without interrupting the normal functioning [3]. As a consequence, interventions are scheduled only if needed. Automated online monitoring, though, should be supported by effective models that can reliably assess aging phenomena. From this perspective, the literature shows that existing solutions still need to be improved. Among the diagnostic techniques, partial discharges (PDs) are a valuable indicator of the insulation condition [4,5]. According to the standard IEC 60270, a partial discharge is a localized, electrical discharge that only partially bridges the insulation between conductors; it can (or cannot) occur adjacent to a conductor. In practice, PDs are both the symptoms and the causes of electrical aging in insulation systems. The phenomenon starts when the electric field locally exceeds the breakdown strength limit of insulating material: a local concentration of electrons yields a polarization that causes an electron avalanche. PDs can lead to the breakdown of the insulation. Thus, by detecting PDs, one can estimate the aging status of the insulation. Accordingly, one can program the corresponding maintenance intervention [6].

The literature proved that online monitoring systems can exploit PD detection to support predictive maintenance. Indeed, this paper aims to address three prominent issues of automated online aging monitoring. The first issue is the ability to infer the lifespan of the apparatus under analysis by exploiting a mathematical model. Most of the existing solutions rely on explicit models, which require a high level of expertise in the system [7]. In fact, one, in general, uses empirical observations to set the parameters of such a mathematical model. Explicit models become impracticable when the complexity of the device under analysis grows, as the challenge is to identify, a priori, a proper function that

fits the relation between inputs and outputs well [8]. In real applications, the monitored apparatus (e.g., industrial motor, oiled-paper insulated transformers, gas-insulated systems, etc.) is compounded by many electrical and mechanical parts, making it troublesome, considering all kinds of factors that can age the insulation during normal functioning. Furthermore, it is very difficult to take into account disruptive changes (e.g., power supply over-voltage), which may accelerate the degradation of the system. Indeed, the explicit models cannot predict online the current status of the insulation system, but they can only state the expected time-to-failure [9,10]. Thus, it seems unrealistic to build an explicit empirical model that can reliably predict the aging status of a complex apparatus. Artificial Intelligence (AI) overcomes this issue by learning models directly from the data (namely, a data-driven approach), without any kind of prior knowledge of the problem. The existing approaches that use AI for the predictive maintenance of insulation systems are categorized as severity techniques and are mostly based on clustering strategies [11–13]. Thus, they assign the apparatus to one among a predefined set of levels, which qualitatively represent the life conditions of the insulation system [14].

The second issue is the selection of the features that will be processed by the model that infers the aging phenomena. First of all, the features must characterize the phenomena. Indeed, aging prediction requires information about the status of the insulation system. A valuable option is the phase-resolved partial discharge (PRPD) pattern [15], also known as the PD pattern: a two-dimensional array containing the occurrences of the PDs' quantized amplitudes with respect to the power supply phase. Skilled human annotators can extract significant information about the aging of an apparatus using PD patterns [16]. In addition, these patterns are often adopted with AI techniques for the severity classification task. Actually, most of the AI techniques cannot process PD patterns directly; hence, expert designers proposed hand-crafted descriptors [14,17]. However, the definition of the feature set is critical and is influenced by the characteristics of the device under analysis, involving cost and technical difficulties. Convolutional Neural Networks (CNNs), conversely, do not require a predefined feature space as input. The CNN architectures are organized as stacked layers that exploit the convolution operation; overall, the layers act as filters that progressively extract meaningful features from the input signal. Actually, the training procedure has the objective of properly tuning such filters; as a major result, the task of feature extraction is implicitly transferred to the CNN. In the application at-hand, CNNs can receive as input the raw PRPD, which in practice is a 2-dimensional signal. Moreover, CNNs have already been successfully employed in pattern recognition problems involving partial discharges [18–20]. In view of these considerations, this paper proposes the use of CNNs to automatically extract the feature set from the data.

PD patterns are also useful for overcoming the last issue: existing models struggle to monitor disruptive events that abruptly change the status of the insulation system. Explicit empirical approaches lead to ideal models that, by construction, miss this kind of phenomena [21]. Severity detection techniques can provide qualitative information, as they can distinguish run-time changes in the lifespan of an apparatus. However, they fail in assessing the actual aging of the insulation system online [14,21–25]. In this paper, the proposed framework assigns, at run-time, an aging score to the apparatus by periodically extracting a PD pattern from the monitored device. Thus, it can detect disruptive events in real-time.

This research shows that CNNs can support an effective model for the real-time assessment of the aging status of electrical insulation systems. The proposed approach inherently addresses the three issues analyzed above. Given the availability of a training set, the inference function can be learned without any prior information about the monitored apparatus. Moreover, CNNs extract significant features from raw data naturally. Finally, once the training phase is concluded, the model infers the aging status of the device at time t^* by using only the single PD pattern extracted at time t^* .

The paper addresses three fundamental aspects for the development of the CNN-based framework: (1) the definition of a proper loss function; (2) the selection of the most

convenient architecture for the CNN; and (3) the techniques to be employed to obtain good accuracy even when exploiting a limited amount of data during the training phase.

The experimental session involved a set of specimens that underwent aging tests according to standard IEC 60851-5. Experimental outcomes proved that the CNN-based framework improved over state-of-the-art techniques in terms of prediction accuracy.

Contribution

This paper shows that CNNs can support an effective methodology to assess insulation lifetime. In the proposed end-to-end approach, a 2-D CNN receives as input PRPD images to estimate the aging status of twisted pairs specimens, thus implementing a regression function. As far as the authors know, this is the first time that an automatic method based on CNN has been employed to predict the remaining life of insulation systems, without exploiting any knowledge of the domain.

Overall, the contribution of the paper can be summarized as follows:

- The use of CNNs in the online assessment of the aging status of electrical insulation systems;
- A design strategy for the effective training of CNNs involving the problem definition, data processing, and model selection;
- An empirical study, with data collected in the laboratory, that confirms the effectiveness of the proposed solution.

The remainder of the paper is organized as follows: Section 2 revises the state-of-the-art; Section 3 introduces the proposed framework and the adopted CNN architecture; Section 4 presents the experimental setup; while Section 5 analyzes the outcomes of the experimental session. Eventually, Section 6 compares the proposal with the state-of-the-art algorithms.

2. Related Works

In the literature, two main categories of works targeted the design of frameworks that assess the status of the insulation system using PDs. The first category addressed the severity classification problem by grouping the PDs based on the condition of the specimen's life status. The severity classification may assess the changes of the apparatus lifespan, but the choice of the number of the classes may yield inaccurate prediction: the higher the number of classes, the higher the number of data required for the training; the lower the number of classes, the higher the risk of classifying apparatuses that can still live for years or are close to breakdown within the same severity category. Indeed, all the methods provide only a qualitative analysis of the aging status of the insulation, without scoring the actual condition. The upper part of Table 1 summarizes the references about severity classification [14,22–25]. The table provides the testing environment, the feature extraction and reduction strategy, and the prediction method in columns 2, 3 and 4, respectively. Finally, a check symbol in the last column distinguishes methods that can detect changes in the insulation system, shortening the life span of the apparatus.

The second category of works builds a life prediction model of aged specimens affected by PDs based on several factors (i.e., power supply voltage and frequency, aging temperature, humidity, pressure, etc.) and their interactions. These works set the influence of each factor on the life model using experimental techniques, such as the design of experiment (DoE) and the response surface method (RSM). Other works used linear regression techniques on features extracted from PRPDs to estimate the lifetime of twisted pairs specimens. All these methods made a strong hypothesis on the relation between the factors and the life duration of the aged specimens. The lower part of Table 1 reports the references for the explicit empirical models [21,26–30]. In [31], a fully unsupervised approach detected changes in the life status of a specimen. The authors showed that such an approach could be combined with the method proposed [29,30] to improve the insulation lifetime predictions. A few drawbacks affect explicit empirical techniques. In fact, these models handle only a limited number of factors, impose strong assumptions on the

relations between the life duration and the factors, and fail at detecting disruptive changes during aging.

In the last years, several studies involved deep neural networks (DNNs) and in particular convolutional neural networks (CNNs), which obtained excellent results in the pattern recognition fields [32–34]. In [35], the authors proved the effectiveness of CNNs for the maintenance of surfaces, predicting pavement cracks in advance. Specifically, in the last 5 years, scientists have employed CNNs to distinguish PD sources [36]. In [37], the authors proposed a framework in which a CNN received as input PRPD images; the framework distinguished six different PD defects created in oil. In [20], a CNN classified PD sources in a gas insulated system (GIS); this approach performed better with respect to the state-of-the-art algorithms. Similarly, other works proved that CNNs obtain an interesting performance in the recognition and classification of PDs [19,38–41].

In the literature, all the methods based on the deep networks identify and classify the defects affecting the insulation systems. Differently, in our proposal, the goal is to adopt CNN to predict the remaining lifetime of the apparatus under monitoring, without explicitly detecting the kind of PDs sources.

Table 1. References table.

Reference	Environment	Feature Extraction and Reduction	Prediction Model	Change Detection
Severity Classification				
[14]	Aged oil–paper insulation system	Statistical parameters from PDs + PCA-based method	Hierarchical clustering	✓
[22]	GIS increasing the tension every 60 minutes	PRPDs analysis and their statistical distributions.	Clustering	✓
[23]	Protrusion defects in GIS	9 features from PRPD distributions	Two-level fuzzy logic	✓
[24]	Aged oil–paper insulation system	PRPD distributions + principle component factor extraction	SVM and Weibull distribution	✓
[25]	Power Transformers	PRDP images	Template matching	✓
Lifetime Assessment				
[26]	Electrical treeing in epoxy resin	PRPD distributions and pulse sequence analysis	Empirical model upon the best factor in terms of monotonicity, prognosability and trendability	×
[21]	Steel plates coated with PEI	Life duration time by varying voltage, frequency, and temperature	Empirical model trained through Design of Experiment (DoE)	×
[27]	Insulation materials of thermal classes 1 and 2	Life duration time by varying voltage, frequency, and temperature	DoE, RSM, Random Tree (RT) and hybrid model (HM)	×
[28]	Three kinds of enamels	Life duration time by varying voltage, frequency and temperature	Lifespan prediction using the Weibull distribution with scatter	×
[29,30]	Twisted pairs specimens	PRPDs features and life duration time with fixed aging factors	Linear regression model	✓

3. CNNs for Aging Assessment

The present paper introduces a model for aging assessment based on CNNs, exploiting the ability of CNNs to deal with complex, non-linear problems when input data can be represented as tensors. A PD pattern represents partial discharges as a 2-dimensional array, that is, a second order tensor. Figure 1 shows how the data collection mechanism is organized. In this plot, the blue line refers to the power supply voltage that sets the reference system for the apparatus under testing; the orange line refers to the partial discharges measured by a suitable sensor. In this example, three occurrences of PDs have been registered; each occurrence is characterized by a pair amplitude–phase. It is worth noting that this pair corresponds to the peak amplitude of a PD.

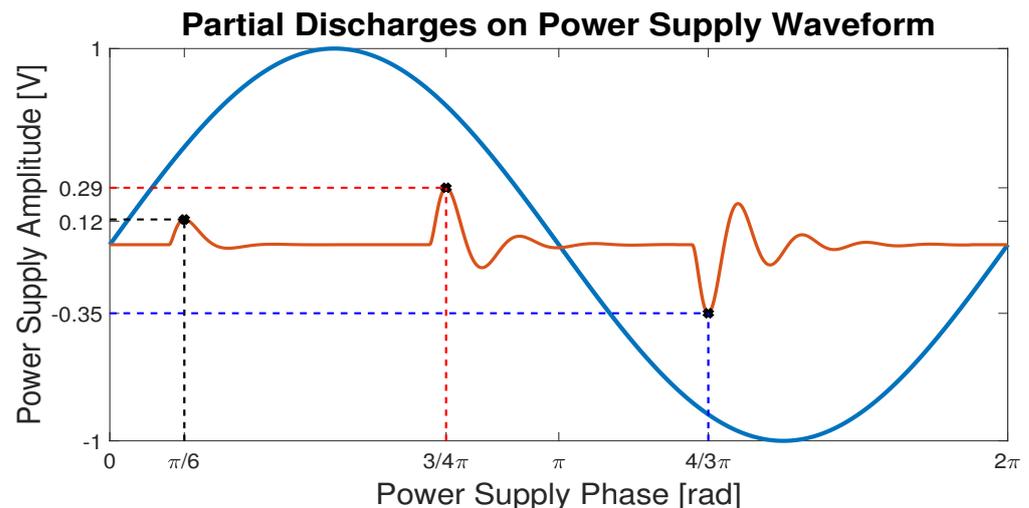


Figure 1. PD impulses over the power supply voltage waveform.

The corresponding PD pattern is organized as a matrix, the columns of which correspond to the power supply phases, and rows mark the maximum amplitudes of the discharges. Thus, each element in the matrix identifies an amplitude–phase pair; cell contents give the occurrences of discharges in a time window T_a . Figure 2 shows an example of a PD pattern.

Accordingly, PD patterns act as inputs for a 2-D CNN designed to infer the aging score of the monitored apparatus. Thus, the CNN supports a regression model that relies on the convolution operation. The regression model is trained by utilizing a proper dataset collected by monitoring a suitable number of apparatuses. Once trained, the model can assess the aging status of new, unseen apparatuses. The Appendix provides details about the general CNN architecture.

The first step in the design of the proposed framework is the definition of the aging score that characterizes the status of the monitored apparatus. The aging phenomena of electrical devices are expected to follow non-linear mechanisms. However, the output of the assessment framework should be analyzed by human users. Hence, a linear aging score seems more informative and easily understandable. Accordingly, an aging score of 0 is assigned to a specimen when the first PD phenomenon arises. An aging score of 0.5 characterizes a specimen at its half-life, while an aging score of 1 is reached when the breakdown is about to occur. This simple rule ensures that the network, when properly trained, will output a score that is user-friendly.

The collection of training data is another fundamental step. A dataset represents a specimen and it is collected from the inception of the first PD until the breakdown, capturing PD patterns at regular steps. Thus, a dataset is an ordered collection of PD patterns. The aging score assigned to the i th PD pattern is:

$$Score_{i_{pattern}} = \frac{t_{i_{pattern}} - t_{start}}{t_{break} - t_{start}}, \quad (1)$$

where t_{start} is the starting acquisition time, t_{break} is the breakdown time, and $t_{ipattern}$ represents the acquisition time of i th PD pattern in the dataset.

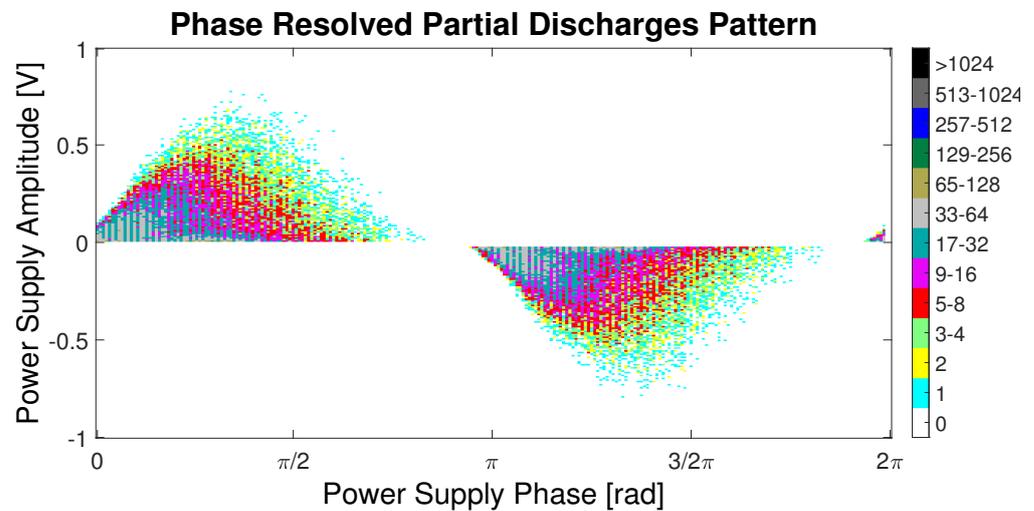


Figure 2. Phase resolved partial discharge pattern representation.

Eventually, the number of datasets matches the number of monitored specimens, and each dataset contains a variable number of PD patterns depending on the lifetime of the specimen, from the PD inception until the breakdown. The union of all the datasets makes the training set $\mathcal{T} = \{(X, y)_i; X_i \in \mathbb{Z}^{256 \times 256}; y_i \in [0, 1]; i = 1, \dots, N\}$, where X_i is a PD pattern and y_i the corresponding aging score. The cost function supporting the CNN training is the mean absolute error between the score inferred by the trained model, $f(X_i)$, and the reference score y_i :

$$loss = \frac{1}{N} \sum_{i=1}^N |y_i - f(X_i)|. \quad (2)$$

Thus, the loss function can be defined as the difference between the score inferred by the CNN and the score expected by assuming a linear behavior in the aging of the apparatus.

Summarizing the previous points, the proposed framework adopts a CNN to address a regression problem. At run-time, the CNN receives a PD pattern and infers the corresponding aging score of the monitored insulation system. To train this CNN, a set of specimens should be monitored until their breakdown to extract the PD patterns and the corresponding aging scores. These data make up the training set, that is, the ground truth supporting the learning process.

In addition, training involves model selection. In principle, the goal of model selection is to properly tune the parameters that characterize the CNN architecture; this process should lead to the final architecture of the CNN, that is, the architecture supporting aging assessment in the framework. On the one hand, the obvious target of model selection is to find the parameters setting that can lead to high accuracy in the prediction. Nonetheless, other constraints should be considered. As the number of layers grows, the number of parameters to be learned also grows. Moreover, the deeper the architecture, the larger the training set; in fact, one may face convergence problems in the training process if the size of the training set is not commensurate with the number of parameters to be learned. This aspect represents a crucial issue for the envisioned application because building a dataset is time consuming; as mentioned above, each specimen should be monitored from the PD inception until the breakdown. Hence, the admissible ranges for the parameters to be tuned should be set by also taking into account such constraints. In practice, one needs to balance the performance in terms of accuracy and the eventual complexity of the involved CNN. Table 2 summarizes the quantities that were set via model selection, along with the admissible values:

- The depth of the network, that is, the number of convolutional layers. The values ranged from 3 to 6;
- The kernel size, which admitted two options: 4×4 and 8×8 ;
- The number of neurons in the fully connected layer, which in the proposed framework involved a single hidden layer. The search space included three values: 16, 32 and 64.

The eventual architecture was also organized according to a few guidelines. First, each convolutional layer was followed by a non-linearity. Second, an average pooling was always stacked on top of two consecutive pairs (convolutional, non-linearity). Third, the number of kernels in the first convolutional layer was set to four. Indeed, starting from the second convolutional layer, the number of kernels always doubled. Doubling progressively the number of kernels is a common practice in deep learning; actually, each layer is designed to learn filters of increasing complexity. Accordingly, the level of abstraction of the features increases in the last layers of the CNNs. Thus, in an architecture with six convolutional layers one would see the following progression in the number of kernels: 4, 8, 16, 32, 64 and 128. Stride was always set to one.

Table 2. Network configurations involved in model selection.

Parameter	Search Space
Kernel size	4×4 , 8×8
Number of layers	3, 4, 5, 6
Size of fully connected layer	16, 32, 64

4. Experimental Setup

The proposed framework has been tested on a specific scenario: the aging of twisted pair specimens. Low voltage stator windings of electrical machines are realized by means of wires insulated by enamels. Thus, the aging of twisted pair specimens can roughly simulate the turn-to-turn failures that can occur during the normal functioning of a winding motor. As a result, the development of predictive techniques for these specimens can be very useful to support the monitoring of low voltage motors insulated by enamels. Accordingly, all the twisted pair specimens involved in the present experimental session, insulated by conventional polyamide-imide enamel, were prepared according to the EIC 60851-5 standard. The following sections explain how training data were collected and how the training process was conducted.

4.1. Data Acquisition

Figure 3 sketches the measurement system adopted to acquire data from a given specimen. A power supply generator provided a sinusoidal waveform with the frequency set to 500 Hz. PD signals affecting the specimen were measured via an HCFT sensor placed around the ground cable, with a band-pass behavior in the range [3 20] MHz. A Picoscope with a resolution of 8 bit, a maximum sampling frequency of 1 GSamples/s and a bandwidth of [0 200] MHz, sampled the PD signals and the power supply voltage. An oven sets the temperature to 150 °C, limiting the impact of environmental effects on the tests (i.e., humidity, non constant temperature, dust, etc.). For each specimen, the peak-to-peak voltage was kept constant. The experiment started at the inception of the first PD and ended at the breakdown of the specimen itself.

The experiments involved different settings for the voltage, as general aging phenomena depends on this parameter [21,27–30]. In particular, aging is faster as the voltage increases. A total of 9 specimens were tested, with voltages ranging between 2000 V and 4000 V. In each experiment, the signals were sampled at regular steps Δ . The amplitude of the input voltage determined the value of Δ , which shortened as voltage increased. Table 3 summarizes the acquisition features of the nine tests. The first column gives the supply voltage; for each row, the table provides the number of specimens tested, the number

of PD patterns collected for each specimen, the value of Δ , and the acquisition time T_a utilized for extracting a PD pattern. Figure 4 shows three PD patterns of a twisted pair specimen aged with a supply voltage of 2000 V. The PD patterns have been acquired in three different moments of the specimen lifespan: at the beginning of the PD activity (score = 0), at the half of the life (score = 0.5), and when the breakdown occurred (score = 1). During the aging of this kind of specimens, the partial discharges tend to increase their number, diminishing the amplitudes by taking into consideration the same acquisition time T_a . For each specimen, the whole PD pattern activity has been acquired from the beginning of the aging phenomenon until the disruptive discharge.

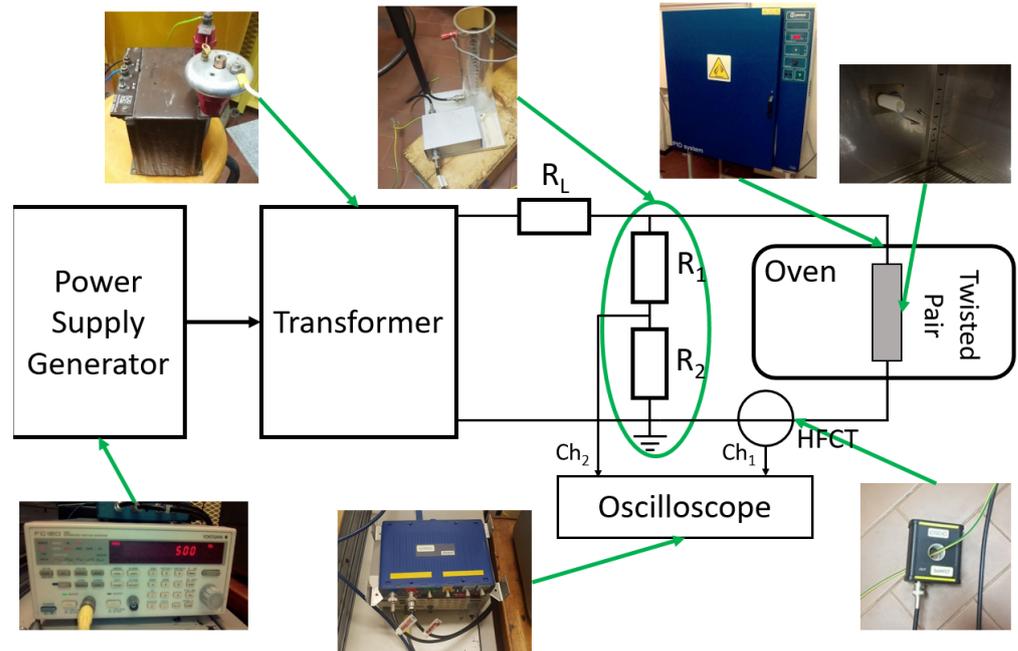


Figure 3. Measurement system.

Table 3. Sub-datasets configurations.

Supply Voltages [V]	Specimens	Num	Δ [sec]	T_a [sec]
2000	2	61	900	40
		73		
2500	3	47	900	10
		56		
		71		
3000	1	52	600	10
3500	2	46	600	10
		51		
4000	1	78	300	10

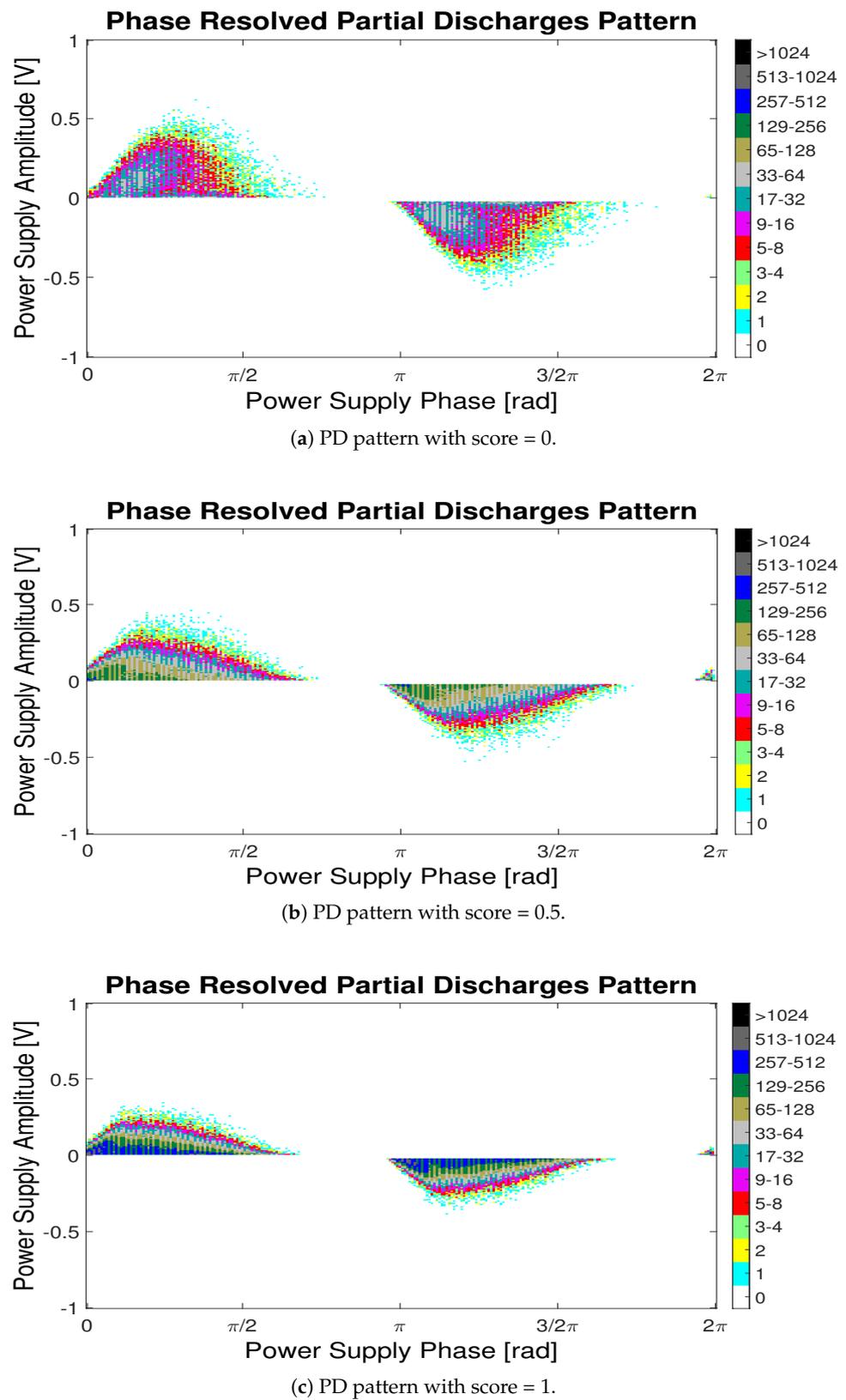


Figure 4. PD Patterns acquired from a specimen aged with a supply voltage of 2000 V at (a) the PD activity inception time (score = 0), (b) half of the life time (score = 0.5), and (c) at the breakdown time (score = 1).

4.2. Training and Model Selection

The whole setup of the training and model selection took into consideration a major constraint: the limited availability of data acquired by monitoring the entire lifespan of a specimen. Such a process is usually expensive and time-consuming. Thus, it is reasonable to assume that one can only exploit a very small dataset. This constraint was indeed taken into account in the setup of the CNN architecture, as discussed in Section 3. Actually, the present framework is designed to rely on a CNN characterized by a limited number of parameters just to avoid convergence issues in the training process.

In the present case, a total of nine specimens with as many experiments were available (as per Table 3): $\mathcal{T}_j = \{(\mathbf{X}, y)_i; \mathbf{X}_i \in \mathbb{Z}^{256 \times 256}; y_i \in [0, 1]; i = 1, \dots, N_j\}$ with $j = 1, \dots, 9$. The available dataset has been split into two non overlapping subsets, that is, a training set and a development set. The training set \mathcal{T}_T included the data collected by seven out of the nine experiments. The development set \mathcal{T}_D included the data collected by the remaining two experiments. This setup was adopted because consecutive PD patterns collected from the same specimen are expected to be strongly correlated.

Model selection was implemented according to the standard hold-out procedure. Thus, for each of the 24 admissible architectures resulting from the search space of Table 2:

1. The learning procedure was completed by using the training set \mathcal{T}_T ;
2. The generalization performance was estimated by computing the mean absolute error (2) of the development set \mathcal{T}_D .

The architecture leading to the lowest mean absolute error was selected for the implementation of the eventual regression model. Algorithm 1 formalizes the steps.

Algorithm 1 Model selection.

Input

- Training set \mathcal{T}_T and development set \mathcal{T}_D .
- Twenty-four architectures ($N_A = 24$), according to Table 2.

1. Learning and selection

for $k = 1; k \leq N_A; k++$ **do**

Train the k -th architecture with \mathcal{T}_T

Test the k -th architecture with \mathcal{T}_D obtaining a loss score \mathcal{L}_k , according to (2)

end for

2. Output Return the best model configuration

$$\Omega = \underset{k}{\operatorname{argmin}} \mathcal{L}$$

5. Experimental Results

The experimental session aimed to evaluate the ability of the CNN-based model to infer the life status of a specimen. The implementation relied on Keras and Tensorflow.

Procedure 1 was adopted to set the configuration of the CNN architecture. The development set \mathcal{T}_D included data from one experiment at 2500 V and one experiment at 3500 V. The selected configuration Ω corresponded to the following setup: 5 layers, kernel size = 8×8 and 16 neurons in the fully connected layer.

The generalization performance of the model was evaluated by using a leave-one-out procedure. Thus, given the set \mathcal{T}_T , six specimens out of seven were utilized in the learning process, while accuracy was assessed on the remaining specimen. This process was repeated seven times to cover all the possible configurations. Accordingly, in the following, S_j will refer to the vector of aging scores obtained when testing with specimen j a CNN trained with the remaining six specimens. S_j is a vector as it collects the ordered sequence of predicted aging scores from t_{start} (first PD pattern extracted from the specimen

under test) to t_{break} (last PD pattern extracted from the specimen under test). Algorithm 2 formalizes the evaluation process.

Algorithm 2 Evaluation.

Input

- Dataset $\mathcal{T}_T = \{(X, y)_i^{(j)}; X_i \in \mathbb{Z}^{256 \times 256}; y_i \in [0, 1]; j = 1, \dots, 7; i = 1, \dots, N_j\}$
- Model configuration Ω

1. Test the best model

for $j = 1; j \leq 7; j++$ **do**

Train the model with \mathcal{T}_T , excluding the j -th specimen

Test the trained model with the j -th specimen of \mathcal{T}_T

Save the vector of aging scores in S_j

end for

2. Output Return S_j , with $j = 1, \dots, 7$.

The outcomes of the seven experiments are grouped in two figures (Figure 5a,b). Figure 5a refers to the test involving the four specimens aged with a power supply voltage in the range [2000 2500] V. For each specimen, the figures show all the predicted scores S_j , that is, from the PD inception until the breakdown. It is worth stressing that during the online monitoring the CNN outputs only a score at a time, that is, when a PD pattern is collected. The figure shows the normalized aging time on the x-axis; that is, a value of 0.5 indicates that the specimen in that instant reached half of its lifetime. The y-axis shows the aging score inferred by the CNN after processing the PD pattern extracted at that instant: four different markers identify the outcomes of the four different experiments. The black line sets the ideal reference; in principle, at any instant the CNN should infer an aging score that corresponds to the normalized aging time. In general, the CNN proved able to estimate accurately the aging phenomena. Problems arise only in the very first part of the lifetime of the specimen. In fact, fast changes affect the insulation material when the PD inception occurs [30,31]. Hence, one may expect the model to be less accurate in that phase.

Figure 5b refers to the test involving the three specimens aged with a power supply voltage in the range [3000 4000] V. In this case, the level of accuracy reached by the CNN is lower. Actually, such an outcome confirms that aging phenomena significantly changes as voltage increases. Nevertheless, all the predictions show a similar trend, where the aging score increases almost monotonically. Thus, it is still possible to assess the aging status of the specimen. As, in practice, a predictive maintenance system aims to detect the insulation deterioration well before the breakdown, even a less accurate prediction can be useful.

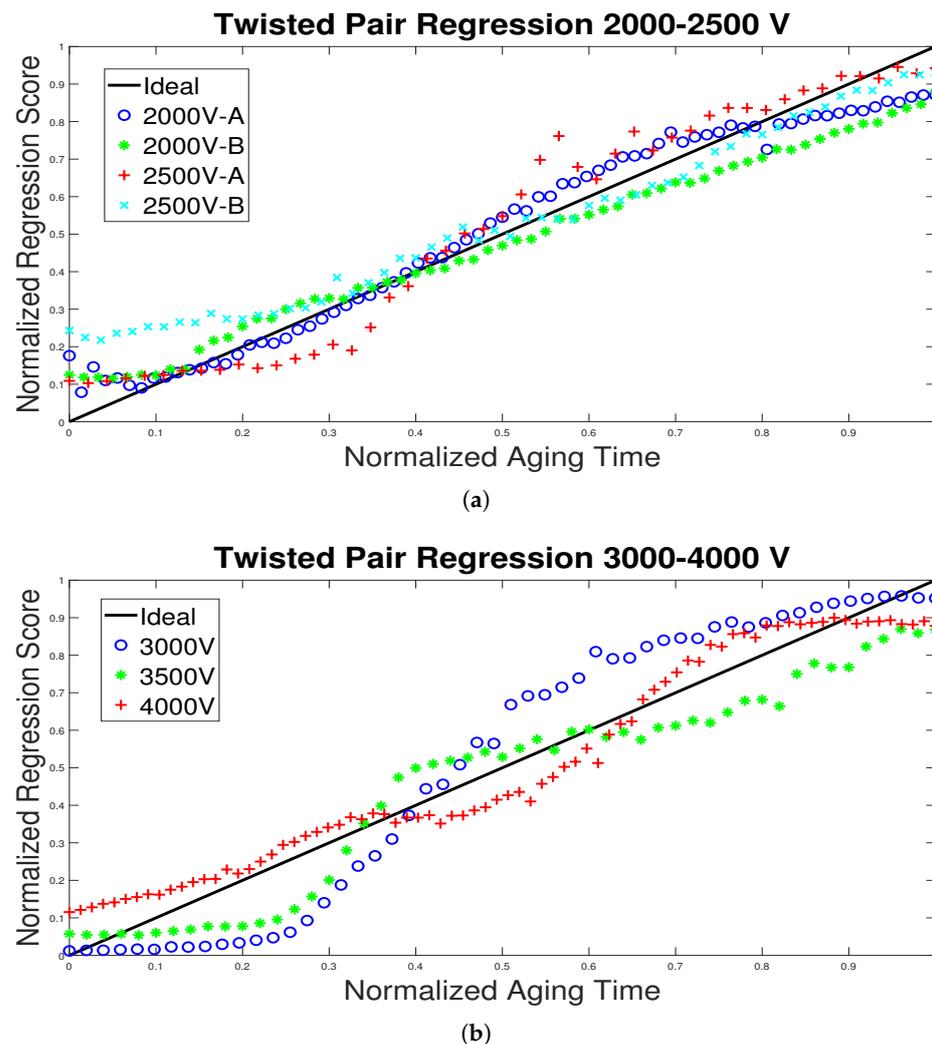


Figure 5. TPs aged with a power supply in the range [2000 2500] V (a) and in the range [3000 4000] V (b).

6. Comparison with State-of-the-Art

This section compares the performance of the proposed CNN-based model with state-of-the-art approaches for aging assessment. The comparison involves (1) ML-based models and (2) explicit models that use empirical observation to set internal parameters.

The ML-based models rely on the approach utilized in designing the proposed framework: the features extracted from the PD pattern feed a regression model implemented by a standard machine learning paradigm. In this case, three different paradigms have been compared: multi-layer Neural Network (NN), linear Support Vector Machine (SVM), and kernel Support Vector Machine (K-SVM). The state-of-the-art provides works that proved the effectiveness of feature extraction techniques for PD pattern classification. In particular:

1. In [42] the features are extracted using Local Binary Pattern (LBP) and Histogram of Oriented Gradient (HOG);
2. In [43], the principal component analysis (PCA) of PD patterns sub-groups based on phase intervals is computed. Besides, the statistical moments (STAT) and the Weibull parameters (WB) from the PD pattern mean-pulse-height distribution and pulse-count distribution are extracted.

Such feature extraction methods can indeed support an aging assessment model. In addition, they proved effectiveness in other applications involving monitoring problems [44,45]. Actually, severity classification methods massively adopt PRPD statistical distributions and techniques based on the PCA to reduce the feature space [14,22–24].

The comparison with explicit models involved the approach presented in [29,30], which shares with the present work the experimental setup (i.e., only the voltage factor influenced the specimens under test, keeping the frequency and the temperature constant). In fact, several papers [21,27,28] affirm that the most influencing factor on the aging condition is the power supply voltage. In the following, Section 6.1 will present the outcomes of the experiments involving approaches based on the pair {feature extractor, ML}, while Section 6.2 will present the outcomes of the experiments involving the model proposed in [29,30].

6.1. Comparison with Approaches Based on ML

The performance of an aging assessment model can be characterized by measuring the absolute error between the predicted aging score \tilde{y}_i at the instant i and the reference score y_i set by the ideal behavior. In fact, one is interested in the distribution of the absolute error over a time window, since the goal is to evaluate the ability of the model to assess aging as the specimen progressively degrades. To the purpose of properly characterizing the performance of an aging assessment model in different conditions, the distribution of the absolute error over a given segment of the specimen lifetime was taken into account. Three segments were considered: the last 25% of the specimen lifespan, the last 50% of the specimen lifespan, and the entire specimen lifespan. Accordingly, the distribution of the absolute error over a given segment was characterized by computing the following quantities:

- MAE: the mean absolute error over a segment

$$MAE = \frac{1}{N - N_0} \sum_{i=N_0}^N |y_i - \tilde{y}_i|; \quad (3)$$

- STD: the standard deviation of the absolute error over a segment

$$STD = \sqrt{\frac{1}{N - N_0} \sum_{i=N_0}^N (y_i - \tilde{y}_i - MAE)^2}. \quad (4)$$

Here, N_0 identifies the starting point of the segment taken into consideration. Thus, for example, to compute MAE and STD for the segment covering the last 50% of the specimen lifespan, one should set $N_0 = N/2$. For the sake of clarity, Figure 6 shows the corresponding configuration. The plot is structured as the plots in Figure 5; thus, the x-axis gives the normalized aging time, while the y-axis gives the aging score inferred by the CNN after processing the PD pattern extracted at that instant. The blue markers identify the outcomes \tilde{y}_i of a predictor; the black line sets the ideal reference y_i . To compute the MAE for the segment involving the last 50% of the lifespan of a specimen, one relies only on the distribution of the absolute errors marked in green.

Table 4 reports the results of the experiments. The first column gives the feature extraction method. The second column identifies the segment of the lifespan utilized for computing MAE and STD. The third column presents the MAE scored by the best predictor over the three considered (NN, SVM, K-SVM). The fourth column displays the difference between the MAE reported in the third column and the corresponding MAE scored by the CNN-based model: a negative value means that the CNN-based model outperformed that predictor. The last column gives the ratio between the STD of the predictor and the STD obtained with the CNN-based model; a value larger than 1 means that the CNN-based model was characterized by a lower STD. Numerical outcomes show that, in general, the CNN-based framework always achieved better performances than the models based on predetermined feature spaces, both in terms of MAE and STD. Figures 7 and 8 provide further details on the outcomes of these experiments. Figure 7 shows the results of the experiments involving a predictor based on linear SVM that processes STAT features; as

per Table 4, this predictor attained interesting performance in terms of MAE over the whole specimens lifespans. These plots are organized as the plots in Figure 5, which showed the results of the experiments involving the CNN. Hence, Figure 7a refers to the test involving the four specimens aged with a power supply voltage in the range [2000 2500] V, while Figure 7b refers to the test involving the three specimens aged with a power supply voltage in the range [3000 4000] V. Figure 7b proves that the predictor processing STAT features could not reliably predict the aging of the specimens under the configuration with a power supply voltage in the range [3000 4000] V. In particular, the unreliable predictions in the second half of the specimens life could be very dangerous, as the apparatus under monitoring is facing the risk of a sudden breakdown.

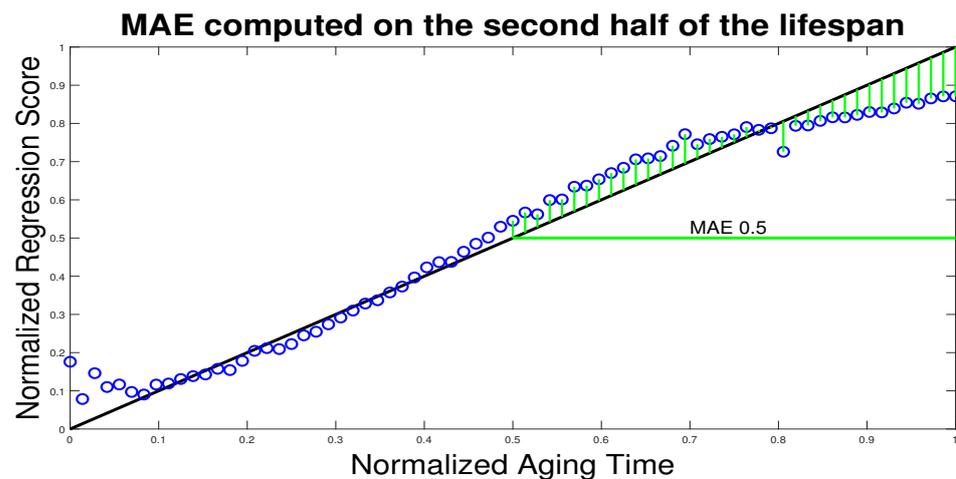


Figure 6. MAE computed over the second half of the specimen lifetime.

Figure 8 shows the results of the experiments involving a predictor based on kernel SVM that processes HOG features; as per Table 4, such predictor also performed effectively in terms of MAE over the whole specimens lifespans. Again, the plots in Figure 8 follow the same structure of the plots in Figure 5. Both Figure 8a,b show that this predictor lacked consistency over the different configurations of power supply voltage. In particular, three configurations proved critical: 2000V – A, 2500V – A, and 3000V.

Table 4. Comparison Results between CNN and AI algorithms.

Features	Segment	MAE	MAE Gap	STD Ratio
LBP	25	0.210	−0.141	3.077
	50	0.159	−0.091	2.820
	100	0.161	−0.097	2.746
HOG	25	0.103	−0.034	1.776
	50	0.113	−0.046	2.003
	100	0.093	−0.030	1.467
STAT + WB	25	0.118	−0.050	3.117
	50	0.092	−0.024	2.414
	100	0.096	−0.033	2.078
PCA	25	0.093	−0.024	1.580
	50	0.083	−0.015	1.429
	100	0.108	−0.044	2.133

6.2. Comparison with the Explicit Empirical Model

The approach presented in [29,30] infers the breakdown time D of an unseen specimen under the hypothesis that voltage is the only aging factor. It relies on two linear

regression models. The first model is entitled to estimate the specimen parameters (Q_0 and I_a , Equation (4) in [29]); the second model computes the multi-linear regression coefficients characterizing the eventual prediction function (K , $n1$ and $n2$, Equation (6) in [29]). Parameters K , $n1$ and $n2$ are tuned by using a training set; however, Q_0 and I_a can be set only by monitoring the specimen itself for a certain amount of time.

According to the leave-one-out strategy (similarly to Procedure 2), six specimens out of the seven included in the training set were utilized to tune K , $n1$ and $n2$. The remaining specimen played the role of the new, unseen apparatus. Thus, Q_0 and I_a were assessed by assuming that a given amount of time could be utilized only to monitor the specimen. In practice, given the specimen under test, D was obtained as follows:

1. Randomly set the starting point of the time window to be used to assess Q_0 and I_a ;
2. Set the length of the time window, collect data and compute Q_0 and I_a ;
3. Compute D .

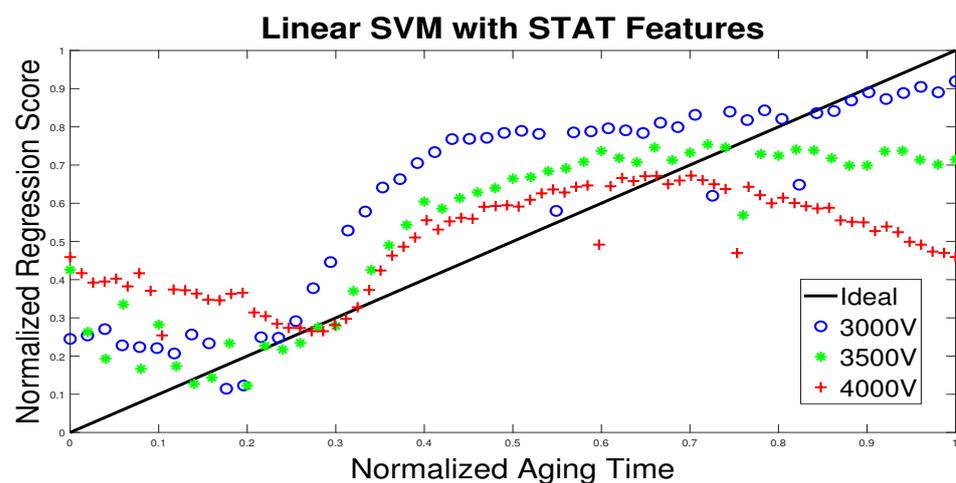
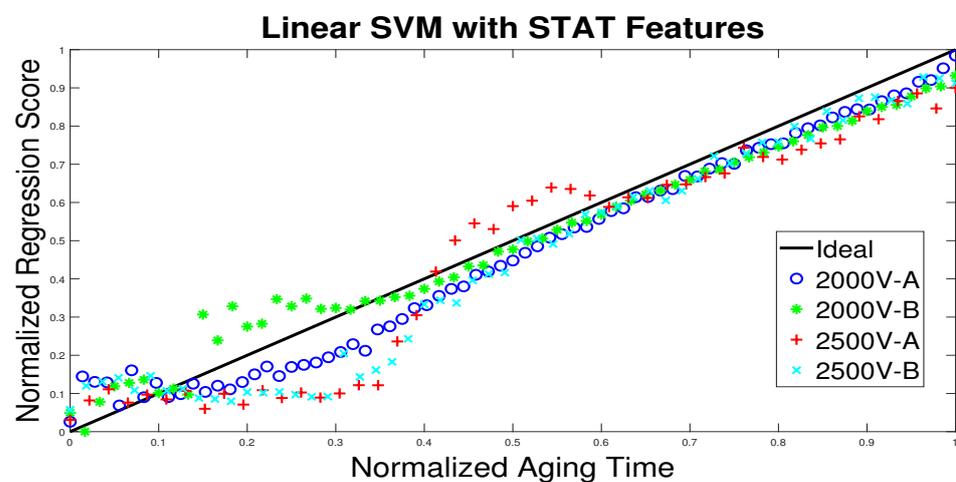


Figure 7. Linear SVM results with STAT features with a power supply in the range [2000 2500] V (a) and in the range [3000 4000] V (b).

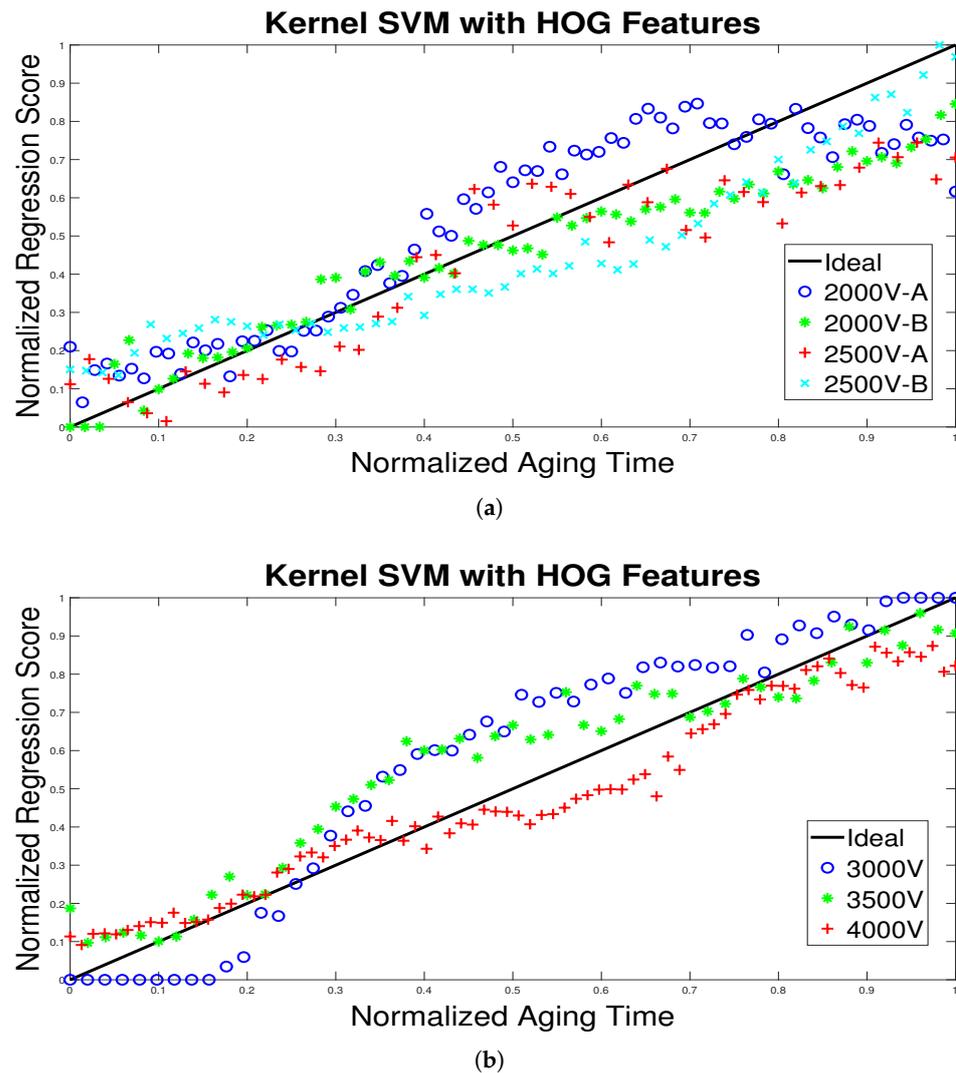


Figure 8. Kernel SVM results with HOG features with a power supply in the range [2000 2500] V (a) and in the range [3000 4000] V (b).

The performance of the empirical model was again estimated by computing the absolute error between the predicted aging behavior (as per D) and the ideal behavior. Three different settings were adopted for the length of the time window: 10%, 25% and 50% of the specimen lifetime. Moreover, the MAE was averaged over 100 different runs, that is, 100 different random starting points.

Table 5 reports the outcomes of the experimental session. The organization of this table is similar to that of Table 4. In the case of Table 5, the first column refers to the length of the time window adopted in the experiments. The remaining columns give the same quantities of Table 4 in the same order.

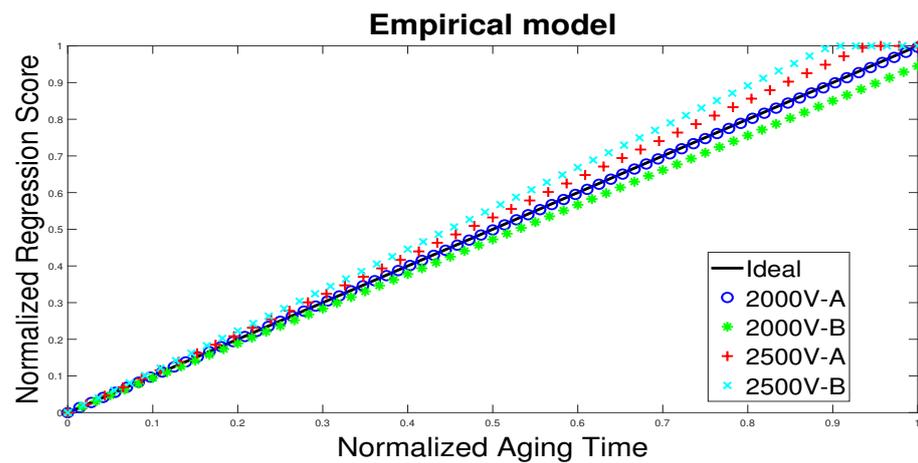
The table shows that the CNN-based framework was also able to outperform the empirical model. It is worth noting, though, that the empirical model attained its best performance when a longer time window was utilized to assess Q_0 and I_a , showing that this approach is not so effective for a real-time application. Such outcome emphasizes the advantages of the proposed architecture in giving a score with only one PD pattern.

Figure 9 shows the outcomes of the experiments involving the empirical model; as per Table 5, the plots refer to the predictor that exploited 50% of time window to assess Q_0 and I_a , since that predictor scored the best MAE. The plots in Figure 9 are structured as in Figure 5. Hence, Figure 9a refers to the test involving the four specimens aged with a power supply voltage in the range [2000 2500] V, while Figure 9b refers to the test

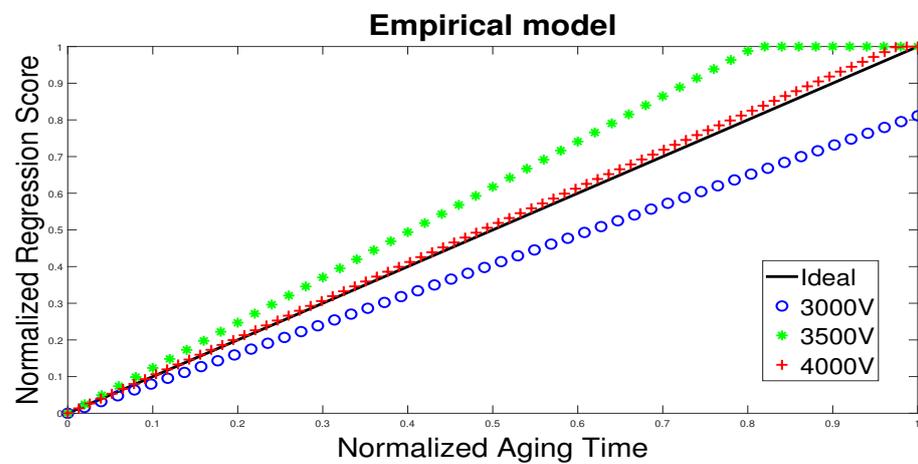
involving the three specimens aged with a power supply voltage in the range [3000 4000] V. As the empirical model relies on a prediction function built via linear regression, the seven predictors (four in Figure 9a and three in Figure 9b) only differ in the pair (Q_0, I_a) . The plots show that the empirical model failed to properly assess the aging process in particular with power supply voltages of 3000 V and 3500 V. In general, by adopting a linear model, one faces the risk of worsening the MAE in the last 25% of the specimen lifespan, which is actually the most critical segment.

Table 5. Comparison Results between CNN and EM.

Time Window	Segment	MAE	MAE Gap	STD Ratio
10%	25	0.172	-0.104	3.256
	50	0.179	-0.111	3.213
	100	0.144	-0.080	3.313
25%	25	0.146	-0.078	2.755
	50	0.160	-0.092	2.726
	100	0.124	-0.060	2.539
50%	25	0.119	-0.051	2.028
	50	0.131	-0.063	2.100
	100	0.105	-0.041	2.216



(a)



(b)

Figure 9. Empirical model with 50% of time window length with a power supply in the range [2000 2500] V (a) and in the range [3000 4000] V (b).

7. Conclusions

This paper presented a novel strategy for the aging prediction of electrical insulation systems. Online aging monitoring is a crucial element for predictive maintenance. Indeed, the literature proves that this task is also very challenging. In several cases, aging prediction is approached as a severity technique, where the goal is to assign the monitored apparatus to a category that qualitatively characterizes the life conditions of the insulation system. The present research, conversely, introduced a framework that—in each instant—can assign an aging score to the insulation system.

The innovative content of the proposed method with respect to state-of-the-art approaches to aging monitoring lies in the ability of learning the feature set starting from a known dataset. In general, state-of-the-art approaches rely on the design of (1) hand-crafted features and (2) an explicit mathematical model that can properly map the feature set into aging scores. Both the tasks, though, involve time-consuming activities and require domain knowledge. In the proposed framework, the aging assessment model relies on PD patterns to obtain information about the partial discharge activity at a given instant. Then, a CNN architecture is demanded to complete feature extraction in the training process. Overall, one exploits the properties of CNNs to avoid issues such as (a) imposing strong assumptions on the relations between the life duration of the apparatus and factors such as power supply voltage, temperature, humidity, and so forth, and (b) modelling disruptive changes during aging.

The experimental activity focused on aging phenomena in twisted pair specimens, which actually can simulate the turn-to-turn failures that occur in winding motors. The CNN-based framework has been compared with two different state-of-the-art approaches:

- ML-based model: ML is exploited to learn the mapping function between a set of hand-crafted features and the aging score;
- Empirical model: the breakdown time of the apparatus is predicted via a multi-linear regression model.

In both cases, experimental outcomes showed the effectiveness of the CNN-based framework, which outperformed the other approaches. The most significant result is the ability of the CNN-based framework to attain consistent performances over different settings of the power supply voltage. This in turn confirmed that a CNN-based approach could better deal with the intricacies of the problem at-hand. The obtained results encourage the adoption of the same approach for more complex insulating systems with the aim of monitoring the related degradation by means of partial discharge measurements.

Author Contributions: The authors contributed as follows: conceptualization, C.G., E.R. and P.G.; data curation, F.G. (Federico Gallesi) and C.G.; formal analysis, C.G. and E.R.; investigation, C.G. and F.G. (Federico Gallesi); methodology, C.G., E.R. and P.G.; project administration, F.G. (Francesco Guastavino); resources, F.G. (Francesco Guastavino) and P.G.; software, C.G. and E.R.; validation, C.G. and E.R.; visualization, all authors; writing—original draft preparation, all authors; writing—review and editing, all authors. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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

Appendix A

CNNs are deep networks that stack multiple layers of convolution, non-linearity and pooling. In the proposed framework, the input \mathcal{I} is a PD pattern. The 2-D CNN receives the tensor \mathcal{I} of dimensions $n \times d$ where n are the rows and d the columns. In general, the architecture of a CNN involves several stacked convolutional layers that differ for the kernel sizes and the number of filters.

A fully connected layer is usually stacked at the top of architecture; this layer is fed with the output of the last convolutional/pooling layer. The fully connected layer is designed to actually implement classification/regression on a set of input features, which in turn are computed by the underlying CNN architecture. In this sense, the CNN plays the role of a feature extractor. The fully connected layer is a standard Artificial Neural Network (ANN) including one or more hidden layers, with neurons embedding non-linear activation functions (e.g., ReLU, sigmoid, tanh). The last hidden layer is directly connected to the output layer, which has as many neurons as the number of classes of the problem at-hand or, in the case of a regression problem, one neuron for each output.

The training procedure of a CNN aims at tuning the parameters characterizing the architecture by exploiting a set of data (i.e., the training set). The amount of parameters stems from the number of convolutional layers, the number of filters and their sizes, and the configuration of the pooling layers. In practice, to train a classification/regression model one should minimize a loss function; usually, in CNNs, stochastic gradient descent is utilized for this purpose. Indeed, standard regularization techniques are employed to avoid over-fitting. The CNN architecture and the optimizer set the computational cost of the training phase. The optimizer determines the number of epochs (i.e., the number of times the entire training set feeds the network), the batch size (sub-set of training data that feeds the network at each training step), the learning rate (i.e., how much update the weights in the gradient descent algorithm). The training procedure tunes the coefficients characterizing the kernels; this in turn means tuning the filters that process input data. Therefore, the CNN architecture is designed to learn the feature extraction process that best models the problem at-hand. In practice, the networks automatically learn the best set of filters.

The architecture of a CNN always involve three core layers:

Convolution. A convolutional layer receives as input a tensor \mathcal{I} of dimensions $n \times d \times c$ where n are the rows, d the columns, and c are the channels. For example, in a RGB image one has $c = 3$, since the image includes a red, green, and blue layer, respectively. The convolution operation applies scalar products between patches of the tensor \mathcal{I} and a kernel tensor \mathcal{K} of dimensions $j \times k \times c$ (also called *filter*), with $j \leq n$ and $k \leq d$; usually $j = k$.

The kernel \mathcal{K} is moved across \mathcal{I} left to right, top to bottom to compute a new tensor \mathcal{O} having dimensions $n' \times d' \times f$ (also called *feature map*), with $n' \leq n$, $d' \leq d$. Both n' and d' depend on the step adopted when shifting the filter \mathcal{K} over \mathcal{I} ; such step is called *stride*. The third dimension f stems from the number of filters applied to tensor \mathcal{I} ; i.e., f is c times the number of filters.

Figure A1a shows the convolution of \mathcal{I} with a kernel \mathcal{K} when $n = d = 4$, $c = 1$, $j=k=2$, and stride equal to 1. Accordingly, the results of the scalar products are stored in the output tensor \mathcal{O} with dimensions $3 \times 3 \times 1$. Figure A1b,c schematize a setup in which multiple filters are applied to \mathcal{I} . As per Figure A1b, the first channel of \mathcal{O} will store the outputs obtained by applying kernel \mathcal{K} . Analogously, as per Figure A1c, the second channel of \mathcal{O} will store the outputs obtained by applying kernel \mathcal{K}' .

Non-linearity. Convolution is a linear operation; thus, a deep network involving only convolutional layers can only model linear functions. Therefore, a convolutional layer is usually followed by a non-linearity; i.e., a non-linear function (also called *activation function*) is applied to the elements of \mathcal{O} . The resulting tensor \mathcal{O}' has the same dimensions of \mathcal{O} . The most adopted activation function in CNNs is the Rectified Linear Unit (ReLU): $f(x) = \max(0, x)$.

Pooling. A pooling layer reduces the dimensions of the input tensor by aggregating local information, producing a tensor \mathcal{O}'' . Basically, the pooling operation is applied after splitting the input tensor in non overlapping patches. For each patch, pooling extracts a scalar value corresponding to the maximum value in a patch (*max pooling*) or to the average value over the patch (*average pooling*). Figure A1d presents an example of max pooling. This operation reduces the tensor size and supports local invariance.

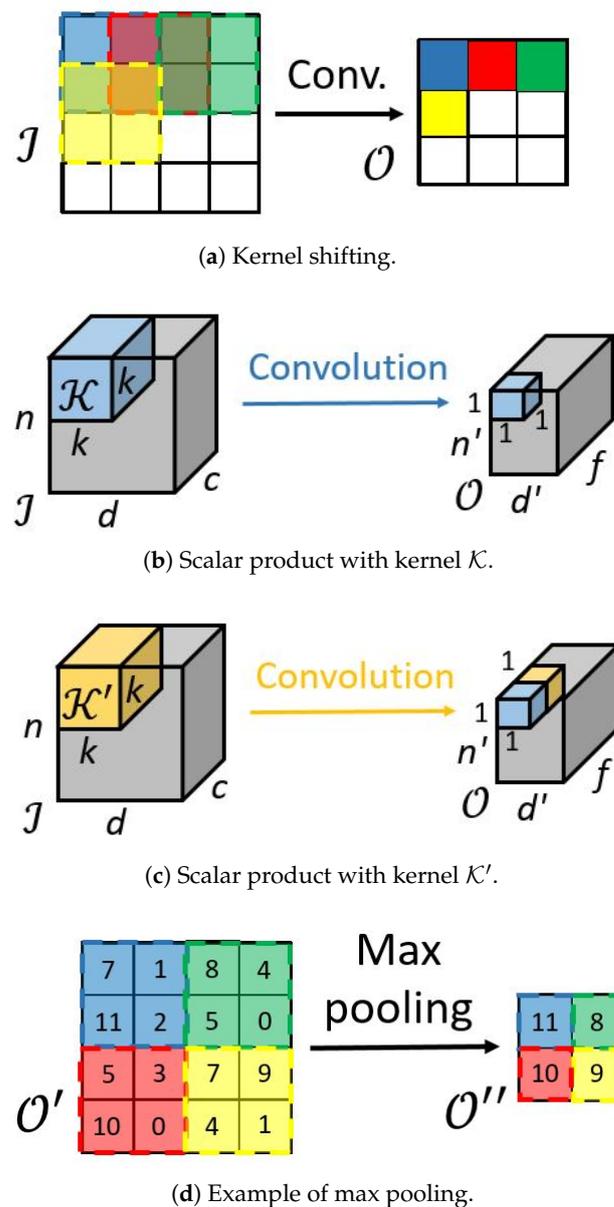


Figure A1. CNN kernel shifting (a), scalar products of the convolution operation (b,c) and max pooling (d).

References

1. Stone, G.C. A perspective on online partial discharge monitoring for assessment of the condition of rotating machine stator winding insulation. *IEEE Electr. Insul. Mag.* **2012**, *28*, 8–13. [[CrossRef](#)]
2. Luo, Y.; Li, Z.; Wang, H. A review of online partial discharge measurement of large generators. *Energies* **2017**, *10*, 1694. [[CrossRef](#)]
3. Guedes, A.S.; Silva, S.M. Insulation protection and online stress agent identification for electric machines using artificial intelligence. *IET Electr. Power Appl.* **2019**, *13*, 559–570. [[CrossRef](#)]
4. Guastavino, F.; Gianoglio, C.; Torello, E.; Cordano, D.; Bruzzone, A. A Predictive Maintenance Remote System based on Partial Discharges Measurements on Wind Turbines. In Proceedings of the 2018 IEEE 2nd International Conference on Dielectrics (ICD), Budapest, Hungary, 1–5 July 2018; pp. 1–3.
5. Morette, N.; Ditchi, T.; Oussar, Y. Feature extraction and ageing state recognition using partial discharges in cables under HVDC. *Electr. Power Syst. Res.* **2020**, *178*, 106053. [[CrossRef](#)]
6. Li, S.; Li, J. Condition monitoring and diagnosis of power equipment: Review and prospective. *High Volt.* **2017**, *2*, 82–91. [[CrossRef](#)]
7. Höpner, V.N.; Wilhelm, V.E. Insulation Life Span of Low-Voltage Electric Motors—A Survey. *Energies* **2021**, *14*, 1738. [[CrossRef](#)]

8. Salameh, F.; Picot, A.; Chabert, M.; Maussion, P. Regression methods for improved lifespan modeling of low voltage machine insulation. *Math. Comput. Simul.* **2017**, *131*, 200–216. [[CrossRef](#)]
9. Kan, M.S.; Tan, A.C.; Mathew, J. A review on prognostic techniques for non-stationary and non-linear rotating systems. *Mech. Syst. Signal Process.* **2015**, *62*, 1–20. [[CrossRef](#)]
10. Li, X.; Ding, Q.; Sun, J.Q. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliab. Eng. Syst. Saf.* **2018**, *172*, 1–11. [[CrossRef](#)]
11. Carvalho, T.P.; Soares, F.A.; Vita, R.; Francisco, R.D.P.; Basto, J.P.; Alcalá, S.G. A systematic literature review of machine learning methods applied to predictive maintenance. *Comput. Ind. Eng.* **2019**, *137*, 106024. [[CrossRef](#)]
12. Lu, S.; Chai, H.; Sahoo, A.; Phung, B. Condition monitoring based on partial discharge diagnostics using machine learning methods: A comprehensive state-of-the-art review. *IEEE Trans. Dielectr. Electr. Insul.* **2020**, *27*, 1861–1888. [[CrossRef](#)]
13. Hassan, W.; Mahmood, F.; Hussain, G.A.; Amin, S. Risk assessment of low voltage motors based on PD measurements and insulation diagnostics. *Measurement* **2021**, *176*, 109151. [[CrossRef](#)]
14. Liao, R.J.; Yang, L.J.; Li, J.; Grzybowski, S. Aging condition assessment of transformer oil-paper insulation model based on partial discharge analysis. *IEEE Trans. Dielectr. Electr. Insul.* **2011**, *18*, 303–311. [[CrossRef](#)]
15. Rudd, S.; McArthur, S.; Judd, M. A generic knowledge-based approach to the analysis of partial discharge data. *IEEE Trans. Dielectr. Electr. Insul.* **2010**, *17*, 149–156. [[CrossRef](#)]
16. Hudon, C.; Bélec, M. The importance of phase resolved partial discharge pattern recognition for on-line generator monitoring. In Proceedings of the Conference Record of the 1998 IEEE International Symposium on Electrical Insulation (Cat. No. 98CH36239), Arlington, VA, USA, 7–10 June 1998; Volume 1, pp. 296–300.
17. Miao, Y.; Jin, M.; Tang, J.; Zeng, F.; Zhou, S.; Zhang, Y. Calculation Method of Partial Discharge Severity Assessment Index Weight Using Factor Analysis Based on Mutual Information. In Proceedings of the 2018 IEEE International Conference on High Voltage Engineering and Application (ICHVE), Athens, Greece, 10–13 September 2018; pp. 1–4.
18. Wang, Y.; Yan, J.; Sun, Q.; Li, J.; Yang, Z. A MobileNets Convolutional Neural Network for GIS Partial Discharge Pattern Recognition in the Ubiquitous Power Internet of Things Context: Optimization, Comparison, and Application. *IEEE Access* **2019**, *7*, 150226–150236. [[CrossRef](#)]
19. Li, G.; Wang, X.; Li, X.; Yang, A.; Rong, M. Partial discharge recognition with a multi-resolution convolutional neural network. *Sensors* **2018**, *18*, 3512. [[CrossRef](#)]
20. Song, H.; Dai, J.; Sheng, G.; Jiang, X. GIS partial discharge pattern recognition via deep convolutional neural network under complex data source. *IEEE Trans. Dielectr. Electr. Insul.* **2018**, *25*, 678–685. [[CrossRef](#)]
21. Lahoud, N.; Faucher, J.; Malec, D.; Maussion, P. Electrical aging of the insulation of low-voltage machines: Model definition and test with the design of experiments. *IEEE Trans. Ind. Electron.* **2013**, *60*, 4147–4155. [[CrossRef](#)]
22. Qi, B.; Li, C.; Xing, Z.; Wei, Z. Partial discharge initiated by free moving metallic particles on GIS insulator surface: Severity diagnosis and assessment. *IEEE Trans. Dielectr. Electr. Insul.* **2014**, *21*, 766–774. [[CrossRef](#)]
23. Zeng, F.; Dong, Y.; Tang, J. Feature extraction and severity assessment of partial discharge under protrusion defect based on fuzzy comprehensive evaluation. *IET Gener. Transm. Distrib.* **2015**, *9*, 2493–2500. [[CrossRef](#)]
24. Xie, P. Analysis of fault of insulation aging of oiled paper of a large-scale power transformer and the prediction of its service life. *IEEJ Trans. Electr. Electron. Eng.* **2019**, *14*, 1139–1144. [[CrossRef](#)]
25. Tenbohlen, S.; Coenen, S.; Djamali, M.; Müller, A.; Samimi, M.H.; Siegel, M. Diagnostic measurements for power transformers. *Energies* **2016**, *9*, 347. [[CrossRef](#)]
26. Aziz, N.; Catterson, V.; Rowland, S.; Bahadoorsingh, S. Analysis of partial discharge features as prognostic indicators of electrical treeing. *IEEE Trans. Dielectr. Electr. Insul.* **2017**, *24*, 129–136. [[CrossRef](#)]
27. Salameh, F.; Picot, A.; Chabert, M.; Maussion, P. Parametric and nonparametric models for lifespan modeling of insulation systems in electrical machines. *IEEE Trans. Ind. Appl.* **2017**, *53*, 3119–3128. [[CrossRef](#)]
28. Szczepanski, M.; Malec, D.; Maussion, P.; Manfé, P. Design of Experiments Predictive Models as a Tool for Lifespan Prediction and Comparison for Enamelled Wires Used in Low-Voltage Inverter-Fed Motors. *IEEE Trans. Ind. Appl.* **2020**, *56*, 3100–3113. [[CrossRef](#)]
29. Guastavino, F.; Dardano, A.; Torello, E. Measuring partial discharges under pulsed voltage conditions. *IEEE Trans. Dielectr. Electr. Insul.* **2008**, *15*, 1640–1648. [[CrossRef](#)]
30. Guastavino, F.; Gianoglio, C.; Torello, E.; Ferraris, M.; Gianelli, W. Electrical Aging Tests on Conventional and Nanofilled Impregnation Resins. In Proceedings of the 2018 IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP), Cancun, Mexico, 21–24 October 2018; pp. 156–158.
31. Gianoglio, C.; Ragusa, E.; Bruzzone, A.; Gastaldo, P.; Zunino, R.; Guastavino, F. Unsupervised Monitoring System for Predictive Maintenance of High Voltage Apparatus. *Energies* **2020**, *13*, 1109. [[CrossRef](#)]
32. Graves, A.; Mohamed, A.R.; Hinton, G. Speech recognition with deep recurrent neural networks. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 6645–6649.
33. Young, T.; Hazarika, D.; Poria, S.; Cambria, E. Recent trends in deep learning based natural language processing. *IEEE Comput. IntelligenCe Mag.* **2018**, *13*, 55–75. [[CrossRef](#)]
34. Ragusa, E.; Cambria, E.; Zunino, R.; Gastaldo, P. A Survey on Deep Learning in Image Polarity Detection: Balancing Generalization Performances and Computational Costs. *Electronics* **2019**, *8*, 783. [[CrossRef](#)]

35. Chen, C.; Seo, H.; Zhao, Y. A novel pavement transverse cracks detection model using WT-CNN and STFT-CNN for smartphone data analysis. *Int. J. Pavement Eng.* **2021**, *2021*, 1–13.
36. Barrios, S.; Buldain, D.; Comech, M.P.; Gilbert, I.; Orue, I. Partial discharge classification using deep learning methods—Survey of recent progress. *Energies* **2019**, *12*, 2485. [[CrossRef](#)]
37. Catterson, V.; Sheng, B. Deep neural networks for understanding and diagnosing partial discharge data. In Proceedings of the 2015 IEEE Electrical Insulation Conference (EIC), Seattle, WA, USA, 7–10 June 2015; pp. 218–221.
38. Dey, D.; Chatterjee, B.; Dalai, S.; Munshi, S.; Chakravorti, S. A deep learning framework using convolution neural network for classification of impulse fault patterns in transformers with increased accuracy. *IEEE Trans. Dielectr. Electr. Insul.* **2017**, *24*, 3894–3897. [[CrossRef](#)]
39. Wan, X.; Song, H.; Luo, L.; Li, Z.; Sheng, G.; Jiang, X. Pattern recognition of partial discharge image based on one-dimensional convolutional neural network. In Proceedings of the 2018 Condition Monitoring and Diagnosis (CMD), Perth, Australia, 23–26 September 2018; pp. 1–4.
40. Duan, L.; Hu, J.; Zhao, G.; Chen, K.; He, J.; Wang, S.X. Identification of partial discharge defects based on deep learning method. *IEEE Trans. Power Deliv.* **2019**, *34*, 1557–1568. [[CrossRef](#)]
41. Karimi, M.; Majidi, M.; MirSaeedi, H.; Arefi, M.M.; Oskuoee, M. A novel application of deep belief networks in learning partial discharge patterns for classifying corona, surface, and internal discharges. *IEEE Trans. Ind. Electron.* **2019**, *67*, 3277–3287. [[CrossRef](#)]
42. Firuzi, K.; Vakilian, M.; Phung, B.T.; Blackburn, T.R. Partial discharges pattern recognition of transformer defect model by LBP & HOG features. *IEEE Trans. Power Deliv.* **2018**, *34*, 542–550.
43. Raymond, W.J.K.; Illias, H.A. High noise tolerance feature extraction for partial discharge classification in XLPE cable joints. *IEEE Trans. Dielectr. Electr. Insul.* **2017**, *24*, 66–74. [[CrossRef](#)]
44. Luo, Y.; Wu, C.M.; Zhang, Y. Facial expression recognition based on fusion feature of PCA and LBP with SVM. *Opt.-Int. J. Light Electron Opt.* **2013**, *124*, 2767–2770. [[CrossRef](#)]
45. Chen, C.; Seo, H.; Jun, C.H.; Zhao, Y. Pavement crack detection and classification based on fusion feature of LBP and PCA with SVM. *Int. J. Pavement Eng.* **2021**, 1–10. [[CrossRef](#)]