Decision tree machine learning pdf download pdf converter software

Department of Electrical & Computer Engineering, University of Manitoba, Winnipeg, MB R3T 5V6, Canada Department of Electrical & Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada Author to whom correspondence should be addressed. Academic Editors: Zhijin Zhang and Hualong Zheng Energies 2022, 15(14), 5005; Received: 14 June 2022 / Revised: 3 July 2022 / Accepted: 5 July 2022 / Published: 8 July 2022 Condition monitoring of high voltage apparatus is of much importance for the maintenance of electric power systems. Whether it is detecting faults or partial discharges that take place in high voltage equipment, or detecting contamination and degradation of outdoor insulators, deep learning which is a branch of machine learning has been extensively investigated. Instead of using hand-crafted manual features as an input for the traditional machine learning algorithms, deep learning algorithms use raw data as the input where the feature extraction stage is integrated in the learning stage, resulting in a more automated process. This is the main advantage of using deep learning instead of traditional machine learning techniques. This paper presents a review of the recent literature on the application of deep learning techniques in monitoring high voltage apparatus such as GIS, transformers, cables, rotating machines, and outdoor insulators. Diagnosis of electrical insulation degradation is essential for monitoring the integrity of an electric power system. A well-known diagnostic method, which has been employed for a number of decades, is the measurement of localized discharges known as partial discharge (PD) [1]. Detecting fault or PD in electric apparatus, such as transformers, rotating machines, cables, gas insulated switchgear (GIS) and outdoor insulators, has always required the knowledge of expertise who are able to characterise and differentiate the different sources of fault, PD, defect, or degradation. Throughout the years, different parameters had to be extracted manually from recorded patterns or signals. The aim has been to use the manually-extracted parameters in order to implement a classifier that would be able to perform the task of differentiation and characterization of fault, PD, defect, or degradation. Though the process is partially automated, the fact that experts have to select the features presented a problem since different features might result in different outcomes. This influences the performance of the classifier due to its dependence on the manually-selected features.Deep learning allows the feature selection stage to be integrated with the learning process, thus making the process all automated. In high voltage (HV) applications, the aim has mostly been to classify or localize faults, defects, or PD that occur in HV apparatus or determine the degradation of insulating material. The abundance of computational capabilities and the existence of big data has allowed researchers in different fields to take advantage of deep learning algorithms. Other than the main purpose of classifying and localizing the PD or fault in HV apparatus, a deep learning algorithm, namely the Generative Adversarial Network (GAN), allows researchers to generate more input data from a limited amount of experimental/simulation results (e.g., see [2]).Classification refers to the process of differentiating between different sources of fault, defect, or PD or levels of degradation. Given that in real life scenarios, fault or PD can happen due to various sources, it is necessary to identify the source. When the source is identified, one can investigate techniques to eliminate that source from the high voltage system. Different sources or causes of fault, defect, PD, or degradation exhibit different characteristics that are unique to each source, making their classification (differentiation) possibly feasible. On the other hand, localization refers to the process of identifying the position of the fault or PD taking place in high voltage apparatus [3].Extensive research has been done on the use of traditional machine learning techniques in high voltage applications, e.g., [4,5,6,7]. This paper only considers the literature that employs deep learning techniques. Furthermore, the focus of this review paper is solely on the application of deep learning in high voltage engineering and not on the deep learning algorithms themselves. Figure 1 gives an overview of the flow of this review paper.The majority of published literature focus on employing deep learning in PD classification and identification. Figure 2 gives an overview of the flow of the research done on the application of deep learning algorithms for classification of PDs in HV apparatuses.A summary of the papers on classification of PD using deep learning is shown in Table 1. Further to PD classification, deep learning has also been used for other applications in transformers and outdoor insulators that are also reviewed in this paper.The organization of the paper is as follows: In Section 2, four deep learning techniques (namely convolutional neural networks, recurrent neural networks, autoencoders, and generative adversarial networks) are briefly reviewed. These are the commonly-used techniques in the HV application. In Section 3, Section 4, Section 5, Section 6, Section 7 and Section 8, a review of papers on the application of deep learning in HV apparatus including GIS, transmission line networks, rotating machines, cables and solid insulators, transformers, and outdoor insulators is presented. Finally, the concluding remarks are presented in the last section of the paper, Section 9.Deep learning is a branch of machine learning that enables data-driven learning of feature representations for input data originating in diverse application domains [43,44,45,46]. Unlike traditional machine learning algorithms, where features need to be extracted explicitly through pre-defined hand-crafted rules, deep learning has the advantage of using raw data, and learn to extract features depending on the task [47]. This is appreciated especially in complex systems where such features are not necessarily known for a given dataset. As a result, deep neural networks subsume the feature extraction step within the learning phase, thereby computing intrinsic representations of the raw input data in an automatic manner.Similar to traditional machine learning, deep learning also has the following three key paradigms: supervised, unsupervised, and reinforcement learning.For the supervised setting, a labeled dataset is required. The type of output can either be continuous (used in a regression problem) or discrete/categorical (used for classification). For unsupervised systems, data with no labels are given, and the objective is either to cluster the data according to their intrinsic characteristics or learn representations which can be later used for downstream supervised or unsupervised settings [48,49,50]. In scenarios involving agent based learning, exhaustive collection of supervised data is often prohibitively difficult. In such situations, reinforcement learning is a powerful paradigm which allows data collection through interaction with the environment [51]. The agent's goal is to learn policies based on the environment in order to maximize long-term expected rewards [51]. In recent years, research in deep reinforcement learning has gained significant traction wherein the agent policies are learnt through deep neural networks [52,53,54] (see Figure 3). Depending on the inputs and the desired outputs for most of the high voltage application, a handful of mainly supervised deep learning algorithms have been of interest in this area of research. In the next section of this review paper, a brief introduction on major supervised deep learning algorithms is presented.Convolutional neural networks (CNNs) represent a class of deep learning architectures which were originally designed for processing data represented in a grid-like topology, e.g., images [47]. A CNN has four main components: convolutional layer, activation function layer, pooling layer, and fully connected layers. Typically, the output of the convolutional layer is passed to an activation function layer where the output of the latter is passed to the pooling layer. In a deep network, this set of the three components are often cascaded multiple times thereby constituting multiple layers and making the network progressively deeper [43,55,56]. While the initial layers usually end up learning low-level features, the deeper layers tend to learn more complex features. The cascade of these layers constitutes the automatic feature extraction stage, and the fully connected layers constitute the classification stage [57]. More details on each of these components are presented below: Convolutional layer: This layer consists of a bank of learnable linear 1D, 2D, or 3D filters, which are also called kernels [58]. In the high voltage applications, usually 1D and 2D CNNs are used. The 1D-CNN, for example, is used with time-series waveforms, whereas in problems involving phase resolved partial discharge (PRPD) patterns or spectrograms, a 2D-CNN is used. Some of the researchers have employed a 2D-CNN for time series waveforms as well, where they considered an image of the signal as an input rather than the 1D data. These filters are convolved with the input data or the output from a previous layer. The output is a set of feature maps, where the number of feature maps is equal to the number of the filters.Activation function layer: The purpose of adding activation layers is to introduce nonlinearity in the input-to-output mapping being learned by the neural network. This is desired because complex data include nonlinear features that need to be detected. Most frequently employed activation functions include sigmoid, ReLU [59] and tanh [60].Pooling layer: The aim of pooling layer is to subsample the output feature maps so that wider receptive fields can be spanned during convolution without increasing the size of the filter kernel. Another advantage of this layer is to provide positional invariance or shift-invariance to the network [61]. Commonly-used pooling operations are maximum pooling and average pooling.Fully connected layers: In a fully connected (FC) layer, every neuron in one layer is connected to every neuron in the next layer. FC layers are also referred to as dense layers in the literature [47]. In a CNN, the input to the first fully connected layer is the output of the last set of the first three components mentioned above, where the corresponding features maps are flattened into 1D vectors. For classification problems, the architecture is appended by FC layers and ends with a classification layer where the number of neurons is equal to the number of classes.A typical CNN architecture is shown in Figure 4. The main advantage of CNNs compared to traditional neural networks is the weight sharing when training the learnable kernels, which reduces the learnable parameters in the network [62].Recurrent Neural Network (RNN) is another family of deep learning architectures which are intended for the processing of sequential data [63,64]. A simple illustration of an RNN model is shown in Figure 5.RNN process data from each time point in a sequential manner. However, the output is not just influenced by data at the current time, but also by the entire history of inputs that have been fed into the RNN previously. This is reflected by the cycles in the architecture, which are maintained in the hidden unit as a state vector including the history of the previous time points. RNN cells have one common set of weights, and when backpropagation runs, data from different time points contribute in updating the same set of weights.The mathematical representation of the RNN is shown as where h t represents the updated state vector, h t − 1 is the hidden state vector from the previous time step, x t represents the input vector at time t, and f w represents a given function corresponding to learnable weight vector w . The input can either be a vector or a sequence, and the output can either be a vector, sequence, or a value. For example, given a high voltage problem where classification of different PD pulses is required, the input is a vector of PD pulses and the output is a label corresponding to a PD source.The drawback of a typical RNN is the long-term dependency where the current state depends on all the previous states, which causes the vanishing gradient problem [47]. The vanishing gradient emerges from the fact that, as RNN processes more time steps, repeated multiplication of small weights causes the gradients to approach zeros. To overcome this problem, long short-term memory (LSTM) architecture is used [65]. The main difference in an LSTM architecture is that, instead of computing the hidden state directly from the previous one, LSTM computes additional states, and this structure allows alternative paths to gradients to flow during the backpropagation avoiding repeated matrix multiplications [66]. An LSTM cell has two hidden states c t corresponding to the cell state and h t corresponding to the hidden state which are calculated as where i is the input gate, f is the forget gate, o is the output gate, and g is the gate. The operator ⊙ is element-wise multiplication operation. The input gate decides what new information will be stored in the cell state, the forget gate decides what information will be removed from the cell state, and the output gate decides what information from the cell state will be used in the output [67].Autoencoders (AE) were introduced in 1980s [68] in order to learn useful representations in an unsupervised fashion by the use of the input data on its own [69]. They were then reintroduced in 2006 with the booming of the deep learning architectures [70]. The idea behind an autoencoder is to train a neural network such that the model learns a latent intrinsic representation of the original input. An autoencoder consists of an encoder–decoder architecture, wherein the role of the encoder is to transform the input to a latent representation, while the decoder is responsible for transforming the latent representation back to the original data. The two parts (encoder and decoder) are learned jointly so as to minimize the reconstruction error between the decoder's output and the network input. A simple illustration of an autoencoder model is shown in Figure 6.Assuming that the input is x and the reconstructed output is x ˜ , the model is trained to minimize the reconstruction error L (x , x ˜) . The encoder and decoder can be fully connected layer networks or any deep learning architecture. The encoder is expressed as a function G such that where b i represents the latent feature representation (bottleneck) of a single observation sample x i . The decoder F accepts b i as input and produces x ˜ i is. This is shown in The goal is then to find F and G that would minimize where the summation is over all the observations during training.Generative models allow sampling data from the probability distribution of a given data. There is a long tradition of learning data distributions including methods for density estimation [71,72]. For high-dimensional continuous data, classical density estimation techniques become intractable. Generative Adversarial Networks (GANs) enable generative models for data stemming from unknown probability distributions [73]. The key idea is to learn a neural network to map a datapoint sampled from a simple distribution (such as normal) to data from the training data distribution. To assess if the generated data have modeled training distribution, another network, known as the discriminator, is trained to distinguish between the generated examples and examples from the original dataset. The goal of the GAN is to train the generator network such that the best discriminator performs as worse as possible. In other words, when the generator parameters have been updated in a way that it becomes difficult to train a classifier for distinguishing between generated and real samples, it implies that the generator is producing outputs that resemble examples from training data distribution. A simple illustration of a GAN model is shown in Figure 7.Classification of PDs has been a standard procedure in the maintenance of high voltage assets. Various distinctive parameters extracted from PD measurements have been introduced for the PD classification application. Starting with current or voltage signals, time series waveforms have been proven to acquire unique behavior for each source of fault or PD [74]. Phase resolved partial discharge (PRPD) patterns have also been used to differentiate between different PD sources [75].Gas insulated switchgear (GIS) is widely used in the industry [76]. A GIS platform is shown in Figure 8.GIS has its components close to each other which makes the fault occurrence in one component transfer to other components easily. The components of the GIS are power conducting components and the control system. The power conducting components are responsible for ensuring the flow of the electric current in the system, and the control systems work on monitoring the behavior of the conducting components. Thermal, mechanical, and electrical faults comprise the main faults that take place in a GIS platform. This section reviews the application of deep learning to solve the problem of classification of only PDs (electrical faults) in GIS. For the condition monitoring of GIS, different detectors have been used in literature including antennas, coupling capacitors, high frequency current transformer (HFCT), UV sensor, thermal sensor and Rogowski coils.There are different sources of PD that take place in a GIS platform. Whether using PRPD patterns or time-series waveforms, we can divide the literature into the following two main groups.In 2017, the authors of [8] simulated four sources of PDs that take place in a GIS platform. The four sources of PDs are: protrusion, contamination, gap, and particle defects. For each of the sources of PDs, four severity states of the PD were collected: normal state, attention state, serious state, and dangerous state. The authors proposed a stacked sparse autoencoder (AE) model, where the output of the middle layer (bottleneck) of the preceding AE is the input to the next AE. The output of middle layer of the final AE is the input to a softmax layer which decided on the assigned severity level label for each sample. The effect of changing different hyperparameters, such as the number of stacked AE or number of nodes in the middle layer, were examined. The proposed model was compared with support vector machine (SVM), where nine statistical characteristics were extracted from the PRPD patterns.The study reports enhanced average classification accuracy of the PD severity compared to SVM.The authors of [12] published in 2018 collected PRPD data from experimental setup and more than 30 live GIS substations. The five defects studied in this work are floating electrode discharge, surface discharge, corona discharge, insulation void discharge, and free metal particle discharge. A known CNN architecture, LeNet-5, was employed and compared with back propagation neural network (BPNN) and SVM, where statistical features were extracted for the latter two and raw data were fed to the CNN. The statistical features extracted from the PRPD patterns include skewness, steepness, asymmetry and cross correlation coefficient of the PD amplitude and rate in both positive and negative half cycles of the applied voltage. In order to optimize the weights of the CNN, the authors trained an autoencoder and used the weights as an initialization for the CNN training. The authors reported an improved average accuracy of the CNN compared to that of SVM and BPNN.A long short-term memory (LSTM) recurrent neural network (RNN) has been used to classify PRPDs in a GIS in [9]. In this work, four different types of defects have been simulated in a controlled lab environment, where the PRPD patterns have been collected. The PD sources simulated are: protruding electrodes, floating electrodes, free particles, and void defects. In addition to that, the authors simulated noise by using an air purifier and the noise signals were obtained using the external UHF sensor. The authors compared the classification accuracy with SVM and a fully dense artificial neural network. The study reported that, although the proposed model takes more training time, the classification accuracy is superior to the other two machine learning models.In 2019, the authors of [13] simulated PDs using a laboratory setup and collected PD data from a live substation. The four PD sources considered in this work that take place in a GIS platform are: floating electrode defects, metallic protrusion defects, insulation void discharge defects, and free metal particle discharge defects. A variational autoencoder (VAE) was trained to extract the eigenvalues corresponding to the PRPD data. The training set included a mix of both the laboratory and substation data. For the test dataset, a matching algorithm based on cosine distance was used in order to decide to what class the test PRPD belongs to. The proposed method was compared with statistical features, deep belief networks (DBN) [77] and CNNs. The authors reported that the eigenvalues extracted from the VAE feature vector have improved results over the other methods used.Four sources of PD that usually occur in GIS (protruding electrodes, floating electrodes, void defects, and free particles) were simulated in [10]. The authors used a Siamese network where the raw input data are pairs of PRPDs. The motivation behind using Siamese network is that PDs usually result in small datasets. Two identical independent CNN models are trained, and the distance between the embedded features resulting from the two CNN models is calculated. As a result, a decision is made whether the pair belong to the same or a different class. The authors compared their proposed architecture with SVM and a CNN model. They reported that the proposed method performed better compared to the latter two.Ref. [11] presents the simulation of artificial defects that take place in a GIS platform, where the PRPD patterns corresponding to each source of PD were recorded. The four PD sources in this work are: corona, floating electrode, particle, and void. The authors proposed a multi-head self attention LSTM based model for PD (LSANPD) and a self attention based neural network model for PD (SANPD). They compared the classification of the two models with their previous published work which used an LSTM-RNN based model. They reported that SANPD and LSANPD are better in terms of classification accuracy and that SANPD is better than the LSANPD and LSTM-RNN model in terms of complexity.Some researchers have used sensors, such as HFCT or UHF sensors, to record voltage or current waveforms induced by PD. The waveforms are in the time domain and referred to as time-series waveforms. This section reviews the papers that have used time-series waveforms instead of PRPD patterns to train and optimize a deep learning algorithm.In 2018, five sources of PD were generated in a GIS tank model in a laboratory setup that includes a floating electrode, a metal protrusion on the conductor and the tank, surface contamination, and free metal particles [35]. Four planar spiral antennas were installed at different locations on the tank. For each signal collected, the authors calculated three different short time Fourier transform (STFT) by changing the window lengths. The different window lengths correspond to high time resolution, high frequency resolution and medium resolution. The proposed model was a CNN-LSTM based model. The three different STFTs calculated from each signal were used to train three different CNN models, where the three outputs of the CNN models are combined by a fully connected layer. The output of the fully connected layer is the input to the LSTM. Since there are four sensors, the model is comprised of four fully-connected layers which are the input to four separate LSTMs. The outputs of the LSTMs collectively decide on the label of each input sample. The authors compared the model performance with other baseline models and with the case where a single window length was used for the STFT. The model showed improved results compared to the other models.In 2019, the authors of [30] investigated GIS PD data which were collected from laboratory experiments and finite difference time domain simulations. A conditional variational autoencoder (CVAE) was used to generate more training data. A seven-layer CNN model was used for the classification of four different sources of PD, free metal particle, metal tip defects, floating electrode defects, and insulation void. The authors also reported a visualization of the feature maps from the first two convolutional layers. They compared their results with support vector machine (SVM), decision trees (DT), back propagation neural network (BPNN), and a few CNN architecture (LeNet5, AlexNet and VGG16). The proposed CNN model outperformed the above mentioned models.Using a laboratory setup, four defects that take place in medium voltage switchgear were replicated [31]. The four sources of PDs included: cable termination floating earth, earth cable in contact with cable termination insulation, voltage presence indicating systems (VPIS) bushing screen disconnected, and earth grounding spring missing on bus bar connector. The spectrogram of the PD signals collected using a coupling capacitor is generated by applying the continuous wavelet transform (CWT). In addition, spectrograms from noises and other HF signals are generated. The authors proposed a convolutional autoencoder (CAE) that is able to reconstruct the spectrograms of the different sources of PDs and noise. After the CAE is trained, the decoder part of the autoencoder is removed and substituted by a fully connected layer followed by the classification layer. This model is trained using a labeled dataset, where the model is able to output the percentage of belonging of a tested spectrogram to each of the four PD sources and the noise/HF signals classes. The study reported high performance ability of the proposed model.In another work [32], four sources of PDs in a gas-insulated switchgear were simulated in a lab setup. An existing CNN architecture, AlexNet, was the method used in this work where the inputs are the time series waveforms treated as images. The results of the proposed method are compared with the fractal method and mean discharge method. The time series waveforms were transformed into a PRPD plot for the sake of applying the fractal method and mean discharge method. Both the fractal and mean discharge methods provide features which are considered as input to two fully connected neural networks. The study included reporting the average classification accuracy of the three models with different percentage of noise added to the signals. The proposed method showed improved results compared with the other two methods especially with a high noise percentage. In addition, the author reported that the time consumed for PD classification was the least using the proposed CNN-based method.In 2021, four types of PDs that take place in a GIS platform (free metal particle defects, metal tip defects, floating electrode defects and insulation void defects) were experimentally simulated and time-series waveforms corresponding to each type of PD were recorded using UHF sensors (butterfly antennas) [33]. The variability in the dataset was introduced by randomly changing the position of the defect. The deep learning model proposed by the authors was based on a depth-wise CNN model where the convolution is divided into two parts: the first part is composed of convolving one channel at a time with the convolution kernel (i.e., depth-wise convolution) and the second part is to mix the feature map using a 1 × 1 convolution kernel (i.e., point-wise convolution). A generative adversarial network (GAN) was also used in order to generate more data. The proposed model was compared with other CNN-based models such as MobileNetV1, MobileNetV2, Xception, ResNet, and LeNet models. The proposed model reported enhanced classification accuracy compared to the other models. Visualization of the feature maps of some layers is presented as well, which highlight what each layer was capable of learning.In the same year, a research group simulated four different sources of PDs in a GIS controlled lab environment [34]. Varying the defect location for each of the artificial defects ensured the variability in the collected dataset. The authors proposed a 1D-CNN model where a multiple scale convolution kernel is used instead of a single scale convolution kernel. Channel shuffling was used on the outputs of the two feature maps produced by the multiple scale convolutional kernel in order to have a unified feature map. Since having labelled data is time consuming and needs expertise knowledge, the authors proposed domain adversarial transfer strategy (DATS), which is inspired by the GAN. Four different unbalanced datasets were acquired from an actual GIS in order to test the performance of the proposed model. The 1D-CNN that was trained on the experimental data are used to classify the on-site GIS PD data where some data had no labels. For the proposed 1D-CNN model, the authors compared the results of the proposed model with traditional 1D- and 2d-CNN models. With regard to the performance of DATS, the authors compared the results with other transfer learning (TL) techniques such as fine-tuning TL and domain adaptation TL. The study reports enhanced results using the proposed framework. The authors suggested that future work will focus on the automatic optimization of hyperparameters and on trying the platform in an online monitoring system.In [36], the aim of the research work was to investigate transfer learning, especially domain adaptive deep transfer learning (DADTL) CNN, for GIS PD diagnosis. The authors used four different datasets for the training of their model. The datasets consisted of measured and simulated data. Dataset A included field data from three types of fault (rolling element, inner ring, and outer ring). Dataset B corresponded to the GIS PD simulation data using the finite difference time domain (FDTD) technique, where four sources of PDs (metal particle, tip, floating electrode, and insulator air gap defect) were simulated. Dataset C corresponded to a 252 kV GIS experimental platform, where signals were captured corresponding to the four defects mentioned in Dataset B. Finally, dataset D corresponded to the PD samples collected from a provincial power company's GIS failure. By the use of maximum mean discrepancy to minimize the sliced Wasserstein distance (SWD), the authors aimed to ensure that transferable features have minimal discrepancy. The proposed model starts with data pre-processing, where samples from the larger datasets are classified as source domain samples, and samples from smaller dataset are classified as target domain samples. The aim is to minimize the differences between the features learned from both source and target datasets. The authors used residuals units in the CNN based architecture. The authors compared the results with traditional CNNs (LeNet and AlexNet) with the same number of layers. The proposed model reported improved results compared to the other deep learning models especially when the dataset is small.Another work which used simulated data are presented in [37], where a CNN-LSTM network is proposed for the classification of PD sources in a GIS system. The model consists of two blocks of convolutional layers followed by pooling layers. The output of the second block is fed to an LSTM layer which is followed by a fully connected layer and ending with the classification layer. In order to generate the dataset, simulation software XFDTD is used. The four sources of PDs are metal tip defect, insulator air gap defect, floating electrode defect, and free metal particle defect. The authors reported the precision, recall, and F1-score of the four sources of PDs. They compared the performance of the proposed model with other models like SVM, LSTM and CNN. The proposed model reported high average classification accuracy compared to the other models.Research has been focused on using different DL techniques for the purpose of identifying PD sources in a GIS platform varying from autoencoders, CNN, LSTM, or a combination of the above techniques. Authors have compared their proposed models with other DL models or traditional machine learning models. Despite the advantages of using different DL techniques, future work should focus on the quality of the input data with regard to the interference, noise, and the fact that multiple PDs can take place at the same time. The integration of the developed models in real-life systems would present its own challenges especially when it comes to developing industrial standards or regulations, and implementing condition-based maintenance asset management policies depending on the severity of the situation.Transmission line networks are used to enable the long distance transmission of power. A few research groups have developed deep learning models to identify PD from non-PD signals collected from a publicly-available dataset. ENET Centre in Czech Republic developed a meter to measure the voltage signal induced by the stray electric field along covered conductors that contained PD or fault signals. The dataset contains noisy real world measurements from high-frequency voltage sensors, where the objective is to identify damaged three-phase, medium-voltage overhead power lines [78].In 2020, the authors of [39] performed pre-processing of the raw data in order to remove noise and low-frequency components of the signals. The output of this process was the time and frequency representation of the signal by applying short-term Fourier transform. The time and frequency domain positive and negative half cycle signals are the input to the proposed deep learning algorithm. The proposed model is a Dual Cycle-Consistency network. Both time and frequency domain branches consist of three blocks. Each block contains a 2D convolutional layer, a Rectified Linear Unit (ReLU), and a batch normalization layer. The output from block-3 of the time-domain and frequency-domain branches is passed through a global average pooling layer, a shared fully connected layer, and a sigmoid layer. In order to calculate the cycle-consistency loss, the outputs are fed to the dual-domain attention module block (DDAM) for joint learning. The prediction is then based on the weighted average of the output from the fully connected layers and the output from the DDAM block. The results are compared with other models such as Random Forest, Resnet18 + VggNet11, and LSTM. The performance metric used is the Matthews Correlation Coefficient (MCC), in addition to precision, recall, and F1-score. The authors reported better results compared to the other approaches.The authors of [40] developed a model based on time-series decomposition and LSTM for to classify PD from non-PD signals from the same public dataset. Seasonal-Trend decomposition using Loess (STL) was used to decompose each raw signal into three parts: trend, seasonal, and residual. PD is mostly reflected in the residual part. Four different STL modules with different seasonal window lengths were used to generate four different residual components. Feature engineering was then applied on the residual parts where a sequential feature vector is extracted. As a result, many-to-one sequential data are generated and are considered as the input to the long short-term memory network (LSTM) classifier. The proposed model was compared with other classifiers such as fully connected layers, SVM, XGBoost, and Multivariate Logistic Regression (MLR). The proposed model showed enhanced classification accuracy compared to the other models.In 2021, the authors of [38] aimed to classify PD versus no-PD signals using the same publicly-available dataset of damaged power lines as the previous paragraph. The proposed model was a traditional 1D CNN model where a Global Average Pooling (GAP) layer is employed before the fully connected layer. Each sample in the dataset is compromised of voltage of the three phases over one period. For each phase, a highpass filter is used to remove the power frequency, after which a maximum filter is used to extract a set of pulses. Each set of pulses from each phase is the input to the trained 1D CNN. Finally, the decision on the label of the power line is based on the three outputs of each phase. In order to visualize what the model is looking at, in order to decide on the label, a pulse activation map (PAM) was used. The evaluation metrics used are Matthews Correlation Coefficient, precision, recall, and accuracy. The authors compared their results with other publicly reported results where models such as LSTM were used. The proposed model showed enhanced results, and the authors suggested that a larger dataset will be more compatible for hyperparameter tuning.In [41], the authors aimed to classify PD versus no PD using the same dataset as the previous papers, which included the three-phase voltage signals. FFT noise reduction algorithms were used on the raw data. The proposed model was a CNN-LSTM model with an attention layer before the classification layer. Starting with two blocks of convolutional and max-pooling layers, the output is fed to a fully connected layer which is considered as the input to the LSTM layer. The output of the LSTM is the input to an attention layer, where multiplication of the feature vector obtained from the LSTM is done with learnable weight coefficients. The output of the attention layer is fed to to a sigmoid which decides on PD versus no PD label. The performance metrics used are precision, recall, and F1-score. The proposed model is compared with other traditional models such as SVM, CNN, and bidirectional LSTM. The study reported higher average accuracy compared to the other models.The papers discussed in this section used the same publicly-available dataset, where different DL approaches were adopted to detect PD pulses from non-PD. This serves as proof that different DL algorithms can give good results; however, the deployment of such algorithms in real-life systems would definitely give a better perspective of what models to use. This section was also a good example to show how various DL techniques can be evaluated using a common dataset.In rotating machines, voltages are generated due to time-varying magnetic fields, which is the result of the change in the flux [79]. The change in the flux results from the mechanical motion of the rotating machine. Rotating machines consist of stator and rotor structures which are made of thin laminations of electrical steel, insulated from each other in order to reduce losses and prevent discharges and faults to take place. Various types of stress, such as thermal, electrical, ambient, or mechanical stress, can affect the insulation system of rotating machines. Statistical data show PD activities have preceded a large number of stator failures [80] and, as such, PD detection in rotating machines has been attracting attention. Extensive research has been done on using traditional machine learning techniques, such as Naïve Bayes-, SVM-, and kNN-based techniques [81], for rotating machine electrical insulation diagnosis. This section mentions the work done using deep learning for the rotating machine PD diagnosis.The PPRDs of a number of sources of PD found in rotating machines through online PD measurements in hydro-generators operating in real-world conditions have been collected in [24]. The PD sources are: internal void, internal delamination, delamination between conductors and insulation, slot, corona, surface tracking, and gap discharges. The authors proposed a methodology to de-noise the PRPDs first and use an image pre-processing technique to separate different clouds in the PRPD patterns. The output of this stage was denoised sub-PRPDs which represent different sources of PD. Three features which are extracted for each sub-PRPD are the input to four different artificial neural networks (ANN). These different ANNs were incorporated in a hierarchical fashion in order to perform the final classification. The authors reported a good overall classification accuracy for all the PD sources.A framework was proposed using visual data analysis for PD source classification in hydrogenerators with a minimum of labeled data [29]. A convolutional encoder was used to project the PD signals acquired from the generator stators to a 2D-visualization latent space. This serves as a visual aid for the expert to analyze the distribution of the training dataset. After being labeled by the experts, the labeled data are trained by a neural network classifier. Other unlabeled data are tested using the already trained classifier, and if any conflict area appeared on the 2D latent space, the human experts will have to label by conflict area sample data. The new labeled data are then added to the dataset, and this procedure is done in an iterative manner until the area of conflicted data is minimized. This study reported a base that integrates both expert knowledge and the advantages of deep learning in order to have a correctly-labeled dataset of PD sources.Although performing preprocessing of the input data is crucial to help the models learn the intrinsic characteristics of patterns (i.e., denoising in HV applications), future work should focus on applying different denoising techniques. Investigating the effect of different denoising techniques on the effectiveness of the DL models yields better understanding of the learning process. Moreover, it is observed that more research is being directed towards exploring the problem of unlabeled data, which is a crucial step for the deployment of any algorithm in a real-life diagnosis system.Electric power can be transmitted by underground cables or by overhead transmission lines. The main advantage of underground cables compared to overhead lines is the low maintenance cost. This is linked to the fact that overhead lines are exposed to environmental factors such as storms or lightning. An underground cable consists of one or more conductors which are covered with suitable insulation and the external component is the protecting cover [82]. The major disadvantage of using underground cables though is the problem of degradation and failure of the insulation under high voltage stress. Hence, detecting PDs/faults is crucial for assessing the health of the system. This section reviews the application of deep learning to solve the problem of classification of faults/PDs in cable insulation and solid dielectric using PRPD patterns or time-series waveforms as the input to train the classification model.Different sources of PD have been classified in a solid insulation in addition to the prediction of the aging stage of the insulation in [14]. The authors of this paper classified three different sources of PD (corona, surface, and internal discharges) which are simulated in a lab environment. They compared the deep belief network (DBN) output with three other machine learning approaches. The input to the DBN was the raw PRPDs, whereas the inputs to the other approaches were features extracted using statistical and vector-norm-based operators. Classification accuracy is the performance metric used in this study. The authors show that the DBN learns distinguishable features without any pre-processing of the PRPDs.The performance of a CNN model was evaluated on the prediction of the ageing stage of high voltage insulation material using PRPD data [15]. Three classes of start, middle, and end as well as noise/disturbance were defined for the electrical insulation degradation process representing the ageing that occurs in an insulation specimen under electrical stress. Precision, recall, and F1 score were the metrics used for the evaluation of the CNN model. The author reported that the performance is consistent even with changes in the CNN hyper-parameters' values.The effect of noise in PRPD patterns on the classification accuracy of different artificial defects in a 11 kV cross linked polyethylene (XLPE) cable joints has been investigated in [23]. There are a total number of five PD sources considered in this study. After training a CNN architecture using noise-free PRPD patterns, transfer learning was performed where the authors used this model to start training another CNN architecture but this time with noisy PRPDs. The results were compared with those obtained using traditional machine learning classifiers where hand crafted features were extracted. The authors reported that the CNN-based model was able to outperform the models that use manual feature extraction with an increase of 16.9 % in the classification accuracy.In [83], the authors simulated five different defects that take place in a 36 kV cross-linked polyethylene (XLPE) cable terminations such as protrusion, void, and corona discharge. The authors used a commercial PD detector in order to capture the PRPD patterns. A CNN architecture was proposed where the authors investigated the effect of different hyperparameters such as pooling and kernel size on the classification accuracy percentage. The authors compared their results with off-the-shelf CNN architectures such as AlexNet, VGG, ResNet and GoogleNet. They reported higher classification accuracy of their proposed model compared to the other models.In 2019, a traditional CNN model was used to differentiate between synthetic PD pulses in power cables [25]. The variability in the synthetic dataset was introduced by the signal-to-noise ratio (SNR) and the position at which the PD initiated. The model was compared with a support vector machine (SVM), where the study reported enhanced results using the proposed algorithm.In the same year, five types of artificial defects in ethylene-propylene-rubber cables in a high voltage laboratory were collected to generate signals containing PD data [26]. Seventeen features were extracted from the time-series waveforms corresponding to characteristics such as pulse width, rise time, fall time, peak voltage, pulse polarity, mean voltage, and root mean square (RMS) voltage. In addition, 16 wavelet features were extracted from the transient signals using Wavelet Transform. In total, 33 features constituted the input corresponding to each signal to the proposed CNN model. Analysis was performed on the effect of the change in the hyperparameters of the CNN architecture such as the number of layers and the convolution kernel sizes. The results were compared with those obtained using SVM and back propagation neural network (BPNN) models. The study reported better classification accuracy when compared with the other two models.In 2020, four common DC insulation faults were simulated during the operation of XLPE cables that include conductor burrs, external semi-conductive layer residue, internal air gap, and scratch on the insulation surface [27]. A modified Canny edge detection operator was used in order to extract the part of the time series signal which includes the PD. A deep belief network is proposed by the authors where the ADAM optimizer is used. The authors compared their model performance with naive Bayes, K-nearest neighbor (kNN), support vector machine (SVM), and back propagation neural networks (BPNN). The authors reported the confusion matrices of the models and compared the classification accuracy. The proposed model outperformed the other methods, and it was reported that, as the training dataset increases, the classification accuracy increases as well.An ensemble of deep learning algorithms was used to differentiate between PD signals and noise signals in medium voltage power cables [28]. The samples were collected from offline in-service cables. The idea behind using ensemble learning is to allow more than one neural network to make the classification decision. CNN, convolutional RNN, LSTM, and bidirectional LSTM (BILSTM) were used in the ensemble frame, but two of these models were used at a time. In this paper, two scenarios were considered: In one scenario where there is difference in the prediction of a sample between two deep learning (DL) models, a human expert will have to decide on the label of that sample, and in the other, the output from the activation function of two different models were added together. Five different cables were tested on the trained models. Adaption training was done for each of the five cables where the classifier layer is re-trained with the measured calibration pulse specific to each cable. The authors reported the results of the five cables for the two ensemble scenarios and with different binary selection of the above-mentioned DL algorithms. They also reported the results when each DL model is used alone. It was reported that the CNN paired with the BILSTM gave the best results.In Ref. [84], the authors aimed to target the problem of losing the voltage signal information that is used for plotting PRPD patterns. Different datasets consisting of three types of cable joint defects were generated in a lab environment, where the signals were recorded using an oscilloscope. Pulse sequence analysis (PSA) was performed by using the change in the magnitude of the PD pulses resulting in a magnitude difference heat map image that was as the input to the proposed CNN model. Investigation was carried out to optimize the CNN hyperparameters, in addition to investigating the effect of the different image features including the size, type, color, and marker size of the images. The authors compared the accuracy of the model with having PRPDs as input versus the PSA as input. The authors claimed that using PSA instead of PRPD yields a higher classification accuracy.The authors in [85] classified and localized ten faults in an 11-kV, three-phase underground cable consisting of various combinations of phase to phase or to ground faults. The dataset was generated by simulation using PSCAD/EMTDC software with varying different system parameters such as fault inception angle and fault location. Additive Gaussian white noise was added to the signals. The authors proposed a CNN-LSTM architecture along with the application of a sliding window technique. The input to the architecture is the current and voltage signals, and the outputs are the fault location, fault inception time, and fault type. The authors compared their results with other deep learning architectures such as CNN and LSTM. The authors reported better performance for their model compared with other models.A simulation model using Matlab was developed in [86] to generate fault signals of an underground cable distribution system consisting of sixteen cables. The fault types are ground fault, short-circuit fault, or open-circuit fault for each of the three phases of the sixteen cables. The authors proposed a deep belief network for this purpose and they compared the classification accuracy with a shallow neural network. The authors reported better results for the proposed deep learning model.The authors of [87] located aged cable segments in underground power distribution systems labeled as even ageing, uneven ageing, and terminal ageing patterns. The signals were captured using an HFCT. The authors proposed a combined stacked autoencoder and CNN architecture for detecting aged segments. When an aged segment is detected, another CNN model was developed to indicate the location and severity of the aged segment. The authors compared the results with other machine learning models such as support vector regression and deep belief network. The proposed model performed better compared with other models for both detection and localization of the aged segments.Ref. [88] proposed a technique to detect the inception faults that take place in cables. The authors used PSCAD/EMTDC in order to simulate the inception fault signals, over-current disturbance signals, and normal current signals. The authors proposed an architecture that includes a sparse autoencoder followed by a deep belief network. They compared the classification accuracy with support vector machine and K-nearest neighbor, where they reported better results performed by the proposed model.The authors of [89] aimed to detect inception faults as well. They used PSCAD/EMTDC as well, where variability in the generated signals was introduced by changing some parameters in the simulation model such as fault impedance and fault location. The authors proposed restricted Boltzmann machine to compress the signals. This was the input to the stacked autoencoder. The authors investigated the trained model performance with simulated and measured data. In addition, they compared the classification accuracy with other models such as CNN, deep belief network, and random forest. The proposed model outperformed the other models.In [90], the authors detected phase-to-ground faults in a typical 10 kV resonant grounding distribution system, which was simulated using PSCAD/EMTDC. The signals were generated under different fault conditions including different fault locations, different grounding resistances, and different fault initial phase angles. Continuous wavelet transform (CWT) was applied to the signals to generate 2D images. The images were transformed to grey-scale, and this was the input to the CNN. The authors investigated the robustness of the model to different parameters such as interference. The authors compared the results with SVM and adaBoost, and they reported that the proposed model gave better results.Although the investigation of different DL techniques is essential for the sake of completeness of any work, it is observed that having common, publicly-available datasets can help in focusing on the generalization of any developed algorithm. In addition, as mentioned in previous sections, the investigation of the effect of any preprocessing technique is necessary in order to have a deeper understanding of the intrinsic characteristics that the DL model is learning.Power transformers play a significant role in power systems, so any failure in this apparatus may interrupt the power supply and cause outages and loss of profit. A photo of a power transformer is shown in Figure 9. One of the beneficial methods for preventing the failure in the power transformers and raising the reliability of these systems is detecting faults in power transformers accurately and promptly.Whether the target is to classify/localize the sources of PDs/faults taking place in a transformer, or to identify overheat or vibration, deep learning has been used for this purpose. The following section summarizes the literature on the use of deep learning for the transformer application.Four typical transformer insulation defects were simulated in [16] that include metal protrusion, oil paper void, surface discharge, and floating potential defects. The authors developed a CNN-LSTM based model where the input is the PRPD data. They compared the results with a CNN-only model and an LSTM-only model. The evaluation metric used was the classification accuracy where the authors reported that CNN-LSTM has better overall recognition accuracy than CNN and LSTM alone.In 2020, the authors of [17] simulated six types of PDs that take place in power transformers using artificial cells in a laboratory setup. They collected the PRPDs of the six PDs which include protruding electrode, moving particle, floating object, surface discharges, bad contact between windings, and void. In order to reduce the input size of the PRPD, the authors used the phase-amplitude (PA) response that is extracted from PRPDs. The authors proposed a CNN model for classification of PD sources. Comparing the classification accuracy of the proposed architecture versus other machine learning classifiers, such as linear and nonlinear SVM, the authors reported a better performance. They also reported that using the PA response as an input increases the accuracy by 1.46 % compared to using the raw PRPDs as the input to the CNN model.The PRPD data of different PD sources in a transformer were collected in a laboratory-controlled setup and reported in [18]. The PD sources included tip discharge, surface discharge, air gap discharge and suspended discharge. The squeeze-and-excitation (SE) module that is a lightweight attention mechanism and the nonlinear function hard-swish (h-swish) were used in addition to a CNN model in order to decrease the accuracy loss of the model further. The authors performed image pre-processing such as segmentation, binarization and enhancement of the data before feeding it to the training model. They compared the results of their model versus other models such as AlexNet, ResNet-18, and VGG16. They reported enhanced average accuracy versus the other models, in addition to less weight storage and reduction of parameters.In the same year, an investigation of a transformer bushing insulation quality, which was affected by poor drying and impregnation, was reported in [19]. The authors used a simple CNN (i.e., 3300 parameters) for the identification of four types of dry impregnation defects using PRPDs as the input to the proposed CNN. The performance metrics used for the evaluation of the model are the precision rate, recall rate, and F1 score. The authors reported 97.1 % average accuracy rate and indicated that their model can be used for online monitoring as it is a small model.A novel convolutional architecture for single and multiple source PD classification, where the model is trained on single-source PDs, was proposed in [20]. The dataset included PRPDs of single and multiple sources of PD taking place in air, oil, and SF6 which mimic common sources of PD. The six single PD sources of floating electrode in SF6, moving particle in SF6, fixed protrusion in SF6, free particle in transformer oil, needle electrode in transformer oil, and corona in air were simulated in a laboratory setup. The proposed architecture has a convolutional backbone feeding into multiple fully connected neural networks (FCNs). The performance metrics used are the arithmetic mean of recall and precision in addition to the classification accuracy and false negative rate. The authors compared their results with one-versus-all CNN and reported that their model has better results than the traditional single-branch CNN architecture.Adam et al. [21] simulated six artificial PD sources in a controlled lab environment that mimics PDs in a power transformer. The PD sources include two discharge sources in air and four discharge sources in mineral oil. The time at which the discharge takes place, the apparent discharge in pC, and the phase angle are recorded for each PD event. In addition, 100 PD events constituted a sample. Superimposed patterns were created by using the single sources patterns, where 30 different combinations of samples with two class labels are formed. The authors proposed an LSTM model which is able to classify multiple and single sources of PDs, where the training was done just on single sources of PDs. The study reported the multi-label accuracy in addition to the single-label accuracy. The multi-label accuracy is defined as the proportion of the correctly predicted labels to the total number of labels for each sample. The model showed a 99 % average accuracy for single PD sources and 43 % for the average multi-label classification problem.Ref. [22] outlined the same objective as the previous two papers. Single and multiple sources of corona discharge in a controlled lab environment were simulated. The four single sources were: sphere–plane, sphere–sphere, needle–plane, and needle–needle. The three multiple sources were: needle–needle and sphere–sphere, needle–plane and sphere–sphere, and sphere–plane and needle–needle. The PRPD patterns were collected, and pre-processing was performed by filtering discharges that have small magnitude. The input to the deep learning models were greyscale images of 75 by 75 pixels. The classes were labeled from 0 for the first single class to 6 for the double-sourced configuration that is considering the multi-source classes as a new class. The authors proposed an optimized ResNet model which they compared with other DL models such as AlexNet, Inception-V3, residual network (ResNet), and DenseNet. The study reported enhanced classification accuracy and least computational cost.The dissolved gas analysis (DGA) is an established method for detecting internal faults in transformers. In recent years, the application of deep learning in DGA has received more attention. This section reviews the literature where the researchers have used deep learning to improve the accuracy of the DGA.The DGA of insulating oil was conducted for transformer fault diagnosis by Dai et al. in 2017 [91]. To improve the efficiency of diagnosis, the authors proposed a novel transformer fault diagnosis approach based on deep belief networks (DBN), which outperforms power transformer fault diagnosis using support vector machine (SVM), back-propagation neural network (BPNN), and ratio methods. A variety of sources were used to collect the input DGA data, including data provided by the State Grid Corporation of China and previous publications. The proposed model was trained using different combinations of DGA ratios associated with fault patterns (the so-called non-code ratios). The training and testing accuracy of 96.4 % and 95.9 % were observed for the DBN with non-code ratios in this study, respectively.In 2020, a semi-supervised autoencoder with an auxiliary task (SAAT) was introduced by Kim et al. to extract a health feature space for power transformer fault diagnosis considering DGA [92]. The DGA dataset was provided by Korea Electric Power Corporation. The proposed SAAT achieved an accuracy of over 90 % in both fault detection and fault identification. The same group has also developed a framework that bridges Duval's method with a deep neural network (DNN) technique for power transformer fault diagnosis employing DGA [93]. The dataset employed contains 4000 unlabeled and 117 labeled DGA data. The obtained results emphasize the superiority of the proposed method compared with the existing AI-based methods in terms of accuracy.Wu et al. introduced a CNN-LSTM deep parallel diagnostic method for transformer DGA employing its ability to extract nonlinear features [94]. The authors showed that this method has a better anti-interference ability compared to the other techniques studied in the paper. In this parallel CNN-LSTM based diagnostic method, the input was in the form of an image derived from the DGA numerical data. The issue of insufficient data was overcome by using the transfer learning technology. The results obtained in this study indicate that the diagnostic accuracy rate is 96.9 % without complicated feature extraction.In 2021, Taha et al. presented a CNN model to precisely diagnose a variety of transformer faults using DGA data and considering different noise levels [95]. The results obtained from applying the proposed method on 589 dataset samples, collected from utilities and literature with various noise levels up to ± 20 % , indicate that the CNN model with combined input ratios improves the prediction accuracy. The obtained accuracy was compared to traditional machine learning methods as well.To improve fault diagnosis in transformers, Hu et al. [96] proposed a method based on refined deep residual shrinkage network (DRSN). The input dataset was based on the amount of gas in the transformer oil, the temperature data, and the number of collected data points based on the timing sequence. The recognition results indicate that the average accuracy of refined DRSN is around 99.67 % for the training set and 97.82 % for the test set. On average, the proposed method could improve the recognition accuracy by 2 % compared to the existing fault diagnosis methods.A probabilistic neural network (PNN)-based fault diagnosis model was presented by Zhou et al. for power transformers [97]. This model optimizes the smoothing factor of the pattern layer of PNN using an optimization technique, improved gray wolf optimizer (IGWO), to enhance the classification accuracy of the PNN. Different fault types data from a real transformer were collected using smart sensors. The obtained results indicate a high diagnostic accuracy of 99.71 % achieved by the IGWO-PNN model.In 2021, Rucconi et al. [98] analyzed the vibration data measured by transformer sensors, such as accelerometers, installed on the transformer tank. These sensors record time series waveforms to build a dataset. An ensemble of fully-connected, feedforward deep neural networks was employed to classify the transformer winding condition (tight or loose). The robustness of the models was investigated by testing them with data collected by sensors at locations other than those used for training. The authors reported a high accuracy in the results.Li et al. used the frequency response analysis (FRA) for detecting the mechanical defects of power transformers in 2021 [99]. The authors employed a lumped-parameter transformer model since creating actual faults experimentally on a real transformer was not practical. The proposed deep learning approach was based on a decision tree classification model and a fully connected neural network that used the FRA data for training. Fifty-five FRA samples were generated as the input to the proposed model by simulating a variety of transformer fault types and levels. The mean absolute error (MAE) and mean square error (MSE) of the validation set were both at a low level, which were employed to reflect the accuracy of the model.A fault diagnosis technique was proposed in [100] by Hong et al. using the vibration analysis. The vibration samples were collected in more than 100 operating transformers and were divided into the categories of normal, degraded, and anomalous. Next, the vibration monitoring data were converted into an image. A deep learning method based on a CNN was employed to classify the images of various input sizes, which indicated an overall accuracy of 98.3 % .A fault diagnosis method based on a deep learning model was presented by Wang et al. applied on a 110 kV three-phase oil-immersed transformer in 2018 [101]. The model used self-powered radio-frequency identification (RFID) sensors and employed the stacked denoising autoencoder (SDA) to learn features. Based on experimental results, the highest accuracy was achieved by the proposed methods and in the shortest time in comparison with other existing methods.In 2021, a method was proposed by Moradzadeh et al. for analyzing a transformer FRA using image processing and a deep learning method, graph convolutional neural network (CNN) [102]. The obtained results using simulation data indicate that the normal mode of CNN (without considering visual images) and with considering the visual images have an accuracy of 97.28 % and 98.33 % , respectively. Using experimental data, an accuracy of 98.01 % and 100 % was reported, respectively.The authors of [103] proposed a CNN model for the identification and localization of faults in transformer winding. The dataset was collected by generating single/multiple disc-to-disc faults of winding insulation in a transformer model at different winding positions, where the current waveforms were recorded. The faults were generated in an analog model of a 33 kV winding of a 3 MVA transformer. The training dataset was generated using an EMTP (Electromagnetic Transient Programming) based digital model, and the test dataset included the data collected from the analog model of the transformer. The results of the CNN model were compared with other methods such as self-organizing maps, fractal features aided SVM, and wavelet-aided SVM. The CNN showed improved classification results compared to the other methods.In 2019, Duan et al. [104] presented an inter-turn fault diagnosis technique to diagnose 15 types of an inter-turn short circuit fault. A multi-channel signal matrix that contains voltage and current waveforms of a simulated transformer was generated as the input of a deep learning-based model. An autoencoder was employed for feature extraction followed by a classifier consisting of convolutional and pooling layers. The proposed model could achieve a recognition accuracy of 99.5 % .An online continuously-operating fault monitoring system for cast-resin transformers was presented by Fanchiang et al. in [105]. They presented an overheating fault diagnosis approach with a maximum accuracy of 99.95 % . The model used infrared thermography (IRT) images as input provided by a thermal camera monitoring system. The images were used to train a Wasserstein autoencoder reconstruction (WAR) model and a differential image classification (DIC) model to classify a number of faults such as inter-turn short circuit or poor contact of primary and secondary sides.Wu et al. introduced an approach to obtain an optimal identification of the operation state of a converter transformer based on vibration detection technology and a deep belief network optimization algorithm [106]. The fused feature extraction technique considered in this study accurately extracted the eigenvectors of the vibration signals. This deep belief network optimization algorithm has offered a high classification accuracy.Similar to the application of DL in any area, the prerequisite of developing any DL algorithm to assess the health of a transformer is the availability of reliable training data. Some data acquisition techniques for transformers are easy to implement even with the transformer in operation (such as DGA or acoustic signals) while some (such as FRA) require an outage. Equipping transformers with sensors that can collect internal data will enhance the application of DL techniques. Most of the papers in the literature either employ simulated lab data or even numerically-generated data. Training a DL algorithm with such data will make the performance of the algorithm in the field require proper validation. Furthermore, research on the application of DL techniques that employ polarization methods data, such as recovery voltage method (RVM), polarization and depolarization currents (PDC), and frequency-domain dielectric spectroscopy (FDS) will enhance the ageing diagnosis of the transformer dielectric materials.Outdoor insulators play an important role in distribution and transmission overhead lines. They mechanically support the high voltage conductors and electrically insulate the high voltage lines from the grounded tower structure. Although they account for approximately 5–8% of the total capital cost of transmission lines, they are responsible for more than 70 % of power line outages [107]. Therefore, it is crucial to continuously inspect them to avoid any risks of premature failure. A photo of a 220-kV overhead transmission line is shown in Figure 10.Outdoor insulators are classified into ceramic and non-ceramic insulators. Despite the differences in their characteristics, both are prone to aging due to the combined effect of electrical, mechanical, and environmental stresses. The main problems associated with outdoor insulators can be categorized into physical defects and pollution related issues. Physical defects like cracks in any parts of the insulator, air voids in the housing material or in the interface between various insulator materials and metallic sharp edges of insulator fittings can cause localized partial discharge (PD) activities to occur. These discharge activities can contribute to the insulation degradation. On the other hand, the accumulation of pollutants on the surface of insulators in the presence of moisture can reduce the leakage resistance and allow leakage currents (LCs) to flow on the surface. Therefore, heat is dissipated, evaporating part of the moisture, and forming dry bands. Since dry bands possess a relatively higher resistance compared to the wet surfaces, voltage stress will concentrate across these bands resulting in the formation of dry band arcing which may lead to complete flashover. As a result, estimating and forecasting pollution levels is critical for utilities to plan their washing schedule to avoid power interruption. Both the Equivalent Salt Deposit Density (ESDD) and the Non-Soluble Deposit Density (NSDD) are commonly used to assess pollution severity.Nowadays, it is estimated that approximately 150 million ceramic insulators are deployed in North American overhead transmission and distribution networks [108]. A significantly high portion of them had either approached or exceeded their lifetime. As a result, utilities are increasingly favoring defective insulator detection systems that are fast, reliable, and cost-effective. To achieve this objective, it is crucial to select the proper sensors and measurement techniques, as well as to use effective machine/deep

learning tools [109,110].Condition monitoring techniques of outdoor insulators can be classified into intrusive and non-intrusive techniques. Since intrusive techniques are not safe, costly and may require the removal of the insulators from the field for further examination, they are time consuming and are not field inspection friendly. Non-intrusive techniques, on the contrary, are faster methods for assessing the health conditions of outdoor insulators and are therefore more favored in field inspections.One of the most common non-intrusive inspection techniques deployed in the field involves the use of manned helicopters equipped with several sensors (like Cameras, IR Cameras, UV Cameras, etc.) for the purpose of recording inspection data. However, this possesses some risk since helicopters require hovering very closely to the electric power transmission line to obtain a better quality of the inspection data. To address this, several alternative solutions were proposed, and they fall mostly under two main categories: Unmanned Aerial Vehicles (UAVs) [111] and Rolling on Wire robots (ROW) [112]. Most of the work in the literature is moving towards UAVs because they have a slight advantage over ROW in the sense that their design does not need to adapt to different physical structures. However, UAVs also exhibit some restrictions in terms of the flying duration, which introduces some challenges when it is required to inspect a large number of insulators in the same trip. The massive volumes of data acquired throughout the inspection process are examined by an experienced crew of human inspectors, hence, the procedure may be time consuming, and the decisions made can be very subjective. Therefore, it is crucial to fuse artificial intelligence modules with UAVs for faster and better inspection performance.In the recent years, machine leaning methodologies have evolved towards the use of deep learning techniques, which have proved to deliver excellent results for pattern recognition problems in a variety of applications. As a result, several authors have used deep learning models to assess the insulator condition. One of the drawbacks of deep learning models is its requirement of a large amount of training data. Despite this limitation, proper use of augmentation techniques can resolve this issue by altering the existing data to create more data for the model training process.The aim of this section is to review deep learning applications along with non-intrusive condition monitoring techniques to assess both ceramic and non-ceramic insulators. In general, the work in the literature can be classified into two main categories, i.e., using deep learning models to detect different physical defects and/or predict the pollution severity level. Both categories relied primarily on either image processing or radiation-based approaches for classification. In the next sections, each category will be discussed along with the most recent research findings.In ceramic insulators, researchers focused on detecting cracks, broken and missing discs in using UAVs. Most of the proposed methods based on deep learning algorithms share the same concept, i.e., the insulator in each image is located using object detection techniques; then, the defect is identified by a pre-trained deep neural network. The existing deep learning algorithms for object detection can be classified into two main categories: one stage and two stage networks [113]. Two stages networks consist of one stage for object detection and another stage for classification, while one-stage networks are end-to-end methods which can predict the position information and classification probability simultaneously in a rapid manner. Generally, two-stage networks possess a higher detection accuracy compared to one-stage networks; however, they have a relatively lower detection speed and thus may not be the best option for real-time operations. Some examples of two stage networks include: regions with convolutional neural networks (RCNN), Fast R-CNN, Region based fully convolutional neural network (R-FCN) and Mask R-CNN and examples of one-stage networks include algorithms like YOLO and single multi-box detector (SSD).An example of two-stage networks is proposed in [114]. It is based on a novel deep CNN cascading architecture. The cascaded architecture is composed of two networks: the first network is responsible for detecting all the insulators in the images by confining them inside detection boxes and cropping them while the rest of the image is discarded. On the other hand, the second network detects the missing caps from the cropped images. The scarcity of the defective images for training was addressed by different data augmentation methods. The precision and recall of the proposed method were found to be 91% and 96%, respectively.To address the issues of slow detection speeds, the authors of [115] proposed a one-stage network using a YOLOv3 deep learning model to recognize and classify images. Moreover, their proposed system combines deep learning with Internet of Things (IoT) through a Raspberry Pi. The work also considered the motion blur in aerial images by implementing a super resolution CNN to reconstruct the blurry images to a high-resolution image before classification. The results show that the proposed system obtains rapid and high accuracy of 95.6% in the identification and classification of insulators' defects. One of the early signs of surface damage of non-ceramic insulators is the loss of their hydrophobicity. Hence, measuring the hydrophobicity is crucial for assessing the insulator surface condition. According to IEC 62073, there are three methods to estimate the hydrophobicity level of insulators, i.e., the contact angle, surface tension and the spray methods [116]. Among these three methods, the spray method is the one that can be applied in the field. The method involves spraying distilled water on the non-ceramic insulator surface; then, the surface can be classified from HC1 (highly hydrophobic) to HC7 (Highly hydrophilic) as shown in Figure 11. The classes are determined based on the size of the wetted area and the contact angle of the droplets. Unfortunately, the main drawback of this method is the subjectivity of human judgment. To overcome this issue, numerous researchers have proposed digital image processing methods to analyze and quantify the hydrophobicity class.In [118], the spray method was used to generate a huge amount of data images which were fed to a deep convolutional neural network model (AlexNet) for the purpose of wettability classification. Compared to other machine learning algorithms, deep learning overcomes the manual dependency on feature extraction and involves less training time due to the transferred learning approach that was used in the article. The algorithm's performance was very promising when compared to other networks like ResNet50, VGGNet16, VGGNet19 and GoogleNet with an overall accuracy of approximately 96%. However, this method may require the removal of the insulator from the field which can be impractical.To resolve this problem, the authors of [119] proposed a method to detect the hydrophobicity of composite insulators using a UAV technology. The drone is equipped with a camera and water spray device in addition to an embedded artificial intelligence (AI) module for non-intrusive classification. Initially, the You Only Look Once version3 (YOLOv3) is used to locate the wet umbrella skirt area of the composite insulator in the complex aerial image, then VGGNet16 was used to classify the Hydrophobicity of the images. An overall classification accuracy of 92.57% was achieved.Other approaches included using image data and deep learning algorithms to assess the material surface degradation in composite insulators. Tracking and Erosion is one of the irreversible physical defects that occurs in non-ceramic insulators which can lead to insulator failure. In [120,121], the authors used transfer learning to train CNN to estimate the severity of erosion in silicone rubber insulators. The algorithm showed robust performance against different lighting conditions which shows the potential of their proposed model in practical applications.To the best of our knowledge, there seems to be a gap in the literature that involves training deep learning models to detect internal and external physical defects using radiation-based measurements like RF antenna and ultrasonic sensors. All the work that has been done on radiation-based techniques involves the use of feature extraction and machine learning techniques [122,123,124,125,126].Several methods have utilized image processing techniques to classify the pollution severity. In [127], for example, a total of 4500 images of ceramic and silicone rubber post insulators were captured under different surface conditions, i.e., clean dry surface, clean with water droplets, contaminated surfaces with cement, contaminated surfaces with soil, wet surface contaminated with soil and wet surface contaminated with cement. Deep CNNs were employed for classification, and a brute-force model selection was introduced to identify and optimize the structure of the CNN classifiers. It was demonstrated that this model selection has achieved a highly accurate architecture. Furthermore, a complexity reduction technique was then applied to achieve lighter architectures. This considers the potential of implementing the CNN classifier in resource limited embedded devices. The results show that this proposed model reduction technique corresponds to a three times lighter architecture at the expense of a slight reduction in the classification accuracy (6.5% only). This is intended to reduce memory usage and flop counts when implemented using embedded devices.In [128], the pollution severity was estimated using UV images. First, Insulator samples were uniformly contaminated with an ESDD level of 0.1 mg/ cm 2 , 0.2 mg/ cm 2 and 0.4 mg/ cm 2 . After applying the voltage, a UV camera was used to capture the discharge activities on the surface of contaminated porcelain insulators. The images were then preprocessed by first graying the image, then changing the pixels to 0 or 255. When doing so, the light spot becomes white, while the rest of the image becomes black, thus highlighting the regional characteristics of the discharge spot in the image. Finally, CNN was used to evaluate the pollution severity of the insulators. It has been found that there is a positive correlation between the pollution level and the severity of the discharge activities under the same voltage level.Other approaches used deep learning models to estimate the LC and an indirect method to estimate the contamination level. For example, the authors of [129] proposed an online monitoring system that uses real-time weather data to predict and classify the LC using bidirectional long short-term memory (Bi-LSTM) model. The sequential weather data consist of parameters like humidity, temperature, rainfall, dew point, solar illumination, wind speed, air pressure and wind direction. They are measured hourly and transferred to data servers. Besides the meteorological data, the LCs are also measured for the purpose of training and validation of the networks using the current transformer. The LC is classified into one of eight groups (levels): i.e., 100 μA–500 μA, 500 μA–1 mA, 1 mA–5 mA, 5 mA–10 mA, 10 mA–100 mA, 100 mA–1 A, 1 A–10 A, and greater than 10 A. Grid search is used to tune the hyperparameters involved in the Bi-LSTM model. The results show that the model achieved an improvement by 12.8% in accuracy compared to other models like LSTM, GRU and RNN.Seven PD sources pertinent to artificially damaged insulator sheds in a controlled lab experiment were simulated in [42]. The first three sources corresponded to damage in one shed of an HV insulator. The other four correspond to damage in two or all sheds in the HV insulator. For this matter, the CNN was used, however, in order to tune the hyperparameters in the CNN architecture, the authors used Bayesian optimization. To generate the training data, the scalogram pattern of the PD signal was generated and transformed using wavelets. Three different mother wavelets (Morse, Amor, and Bump) were used. In addition, different training optimizers' (including stochastic gradient descent with momentums (SGDM), RMSprop, Adam, and Nadam) were used. The authors compared the Bayesian-CNN (B-CNN) with the traditional CNN with no Bayesian optimization in addition to other off-the-shelf deep learning architectures such as VGG19, Resnet 50, and Googlenet. The average classification accuracy was used as the performance metric. The study reported enhanced results for the B-CNN compared to the other architectures with the Bump mother wavelet. The authors also tested the model on another 15-kV porcelain insulator dataset, and the average classification accuracy showed optimistic results which reflect the generalization capabilities of the B-CNN.In general, the literature shows an inadequate amount of work that has been devoted to train deep network architectures to classify and predict pollution levels non-intrusively. All the work done is either focused on applying deep learning models to intrusive measurement techniques [130,131] or applying classical machine learning using non-intrusive techniques [132,133,134,135]. More research is needed to combine deep learning models with non-intrusive approaches for monitoring, particularly those based on radiation type sensors. Furthermore, the majority of the publications were focused on employing one type of sensor for their diagnostics, although this may yield satisfactory results, but could be further improved using multiple sensors. For example, ultrasonic sensors can detect both low and high frequency surface discharges, but it might be difficult to detect internal discharges. On the other hand, RF antennas can be utilized to detect internal and external high frequency discharges but cannot be used to detect low frequency discharges. Thus, combining ultrasonic sensors and RF antenna can be used to detect and classify a wider range of defects.Monitoring of electrical insulation of high voltage apparatus is crucial for the reliable operation of power systems. Such a high voltage apparatus includes but is not limited to gas-insulated switchgear (GIS), transformers, cables, rotating machines, and outdoor insulators. Extensive research has been done on the classification of sources of partial discharge (PD), detection and localization of faults that take place in such apparatus, and the quality and remaining lifetime of insulating material. Modern techniques have been based on machine learning methods, where the input to such methods is composed of manually-extracted features, i.e., feature extraction has required the intervention of human experts. Deep learning, which is a branch of machine learning, has been used to enhance the performance of PD classification, fault and defect detection, contamination diagnosis of outdoor insulators, etc. This enhancement is attributed to the capability of deep learning techniques to use raw data as the input to the classification model. In other words, instead of using manually-extracted features, raw data such as PRPD patterns, time-series waveforms, or images are used as the input to the deep learning systems. This allows the classification model to be fully automated where the feature extracting stage is integrated into the learning stage.In this article, the potential of applying deep learning in assessing the health conditions of different power system assets is highlighted. The following shortcomings/future needs are identified:Most published research employs training data generated in a laboratory environment or by computer simulation that leads to achieving high classification accuracy. A limitation is always presented when data are collected in a controlled lab environment due to the fact that acquiring real data is expensive, intrusive, and time-consuming. Hence, integrating research work in real online or offline systems is always appreciable in order to incorporate all the uncertainties of such systems in the learning process of deep-learning models. Moreover, future research should focus more on the utilization of the use of a generative adversarial network in order to generate more data that mimic real data instead of using lab data that present its own limitations. In addition, future directions should focus more on the utilization of DL techniques such as one shot learning [136] towards the issue of small datasets, which is a typical restriction in the HV application.Prior knowledge of the defect types and/or knowing the exact location of the defect is far from the reality of the field conditions. Moreover, unknown sources and types of external noise may hinder the deep learning algorithms capabilities to identify and/or localize the defect type. Hence, future research needs to focus more on unsupervised learning when it comes to high voltage applications.More work should focus on the occurrence of multiple, simultaneous PDs or faults. The reason is that, in real-life systems, multiple sources of faults or PDs can take place at the same time. Therefore, more focus should be directed towards this problem.One of the limitations of the reported research is the utilization of single sensors like ultrasonic sensors, RF antenna, or IR camera. It is expected that the use of multiple sensors can improve the overall classification accuracy when sensor fusion is applied and different 1D and/or 2D signals are fed to the deep learning classifiers.Integrating the state-of-the-art deep learning algorithms along with promising technologies like drones can improve the inspection efficiency of outdoor insulation systems. With the current improved computational power of micro-controllers, real-time condition monitoring and diagnostics of different defects are feasible using drones and deep learning algorithms.Future research directions should focus on developing electrical insulation ageing models using DL techniques that employ polarization methods data, such as RVM, PDC, and FDS.Using deep learning techniques in the high voltage application is still in the starting stage. More work should be done on deciding on the best standard to specify the optimal architecture per application. One aspect of this is utilizing the use of already established hyperparameter optimization techniques such as the Bayesian optimization technique. In addition, the industrial deployment of the DL algorithms should be addressed, since this requires a different action for each scenario. If the deployment is taking place on a local server, the aim would be to maximize the performance of the algorithm while taking advantage of high-speed and high-end hardware resources. This is different in case the deployment is to take place on a portable monitoring device, where the restriction of space and speed will be presented.With the emerging of the digital twin technologies, deep learning should be utilized for different digital twins of assets, such as transformers or rotating machines. Digital twins are virtual representations of the interactions and behavior that assets can undergo in the physical world. More information on the application of digital twins in power system assets can be found in [137,138,139]. Conceptualization: S.M., A.L., H.M.T., A.A., A.E.-H., B.K.; formal analysis: S.M., A.L., H.M.T., A.A., A.E.-H., B.K.; funding acquisition: A.E.-H., B.K.; investigation: S.M., A.L., H.M.T., A.A., A.E.-H., B.K.; supervision: A.A., A.E.-H., B.K.; Writing—original draft: S.M., A.L., H.M.T.; writing—review and editing: S.M., A.L., H.M.T., A.A., A.E.-H., B.K. All authors have read and agreed to the published version of the manuscript.Financial support from the Natural Sciences and Engineering Research Council of Canada (NSERC) for both the University of Manitoba and University of Waterloo, and the Faculty of Graduate Studies, University of Manitoba is acknowledged.The authors declare no conflict of interest.

Luo, Y.; Li, Z.; Wang, H. A review of online partial discharge measurement of large generators. Energies 2017, 10, 1694. [Google Scholar] [CrossRef]Ardila-Rey, J.A.; Ortiz, J.E.; Creixell, W.; Muhammad-Sukki, F.; Bani, N.A. Artificial generation of partial discharge sources through an algorithm based on deep convolutional generative adversarial networks. IEEE Access 2020, 8, 24561–24575. [Google Scholar] [CrossRef]Sinaga, H.; Phung, B.; Blackburn, T. Partial discharge localization in transformers using UHF detection method. IEEE Trans. Dielectr. Electr. Insul. 2012, 19, 1891–1900. [Google Scholar] [CrossRef]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. [Google Scholar] [CrossRef]Long, J.; Wang, X.; Zhou, W.; Zhang, J.; Dai, D.; Zhu, G. A comprehensive review of signal processing and machine learning technologies for UHF PD detection and diagnosis (I): Preprocessing and Localization Approaches. IEEE Access 2021, 9, 69876–69904. [Google Scholar] [CrossRef]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. [Google Scholar] [CrossRef]Hussain, M.R.; Refaat, S.S.; Abu-Rub, H. Overview and partial discharge analysis of power transformers: A literature review. IEEE Access 2021, 9, 64587–64605. [Google Scholar] [CrossRef]Tang, J.; Jin, M.; Zeng, F.; Zhang, X.; Huang, R. Assessment of PD severity in gas-insulated switchgear with an SSAE. IET Sci. Meas. Technol. 2017, 11, 423–430. [Google Scholar] [CrossRef]Nguyen, M.T.; Nguyen, V.H.; Yun, S.J.; Kim, Y.H. Recurrent neural network for partial discharge diagnosis in gas-insulated switchgear. Energies 2018, 11, 1202. [Google Scholar] [CrossRef]Tuyet-Doan, V.N.; Do, T.D.; Tran-Thi, N.D.; Youn, Y.W.; Kim, Y.H. One-Shot learning for partial discharge diagnosis using ultra-high-frequency sensor in gas-insulated switchgear. Sensors 2020, 20, 5562. [Google Scholar] [CrossRef]Tuyet-Doan, V.N.; Nguyen, T.T.; Nguyen, M.T.; Lee, J.H.; Kim, Y.H. Self-attention network for partial-discharge diagnosis in gas-insulated switchgear. Energies 2020, 13, 2102. [Google Scholar] [CrossRef]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. [Google Scholar] [CrossRef]Dai, J.; Teng, Y.; Zhang, Z.; Yu, Z.; Sheng, G.; Jiang, X. Partial discharge data matching method for GIS case-based reasoning. Energies 2019, 12, 3677. [Google Scholar] [CrossRef]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. [Google Scholar] [CrossRef]Florkowski, M. Classification of partial discharge images using deep convolutional neural networks. Energies 2020, 13, 5496. [Google Scholar] [CrossRef]Zhou, X.; Wu, X.; Ding, P.; Li, X.; He, N.; Zhang, G.; Zhang, X. Research on transformer partial discharge UHF pattern recognition based on CNN-lSTM. Energies 2019, 13, 61. [Google Scholar] [CrossRef]Do, T.D.; Tuyet-Doan, V.N.; Cho, Y.S.; Sun, J.H.; Kim, Y.H. Convolutional-neural-network-based partial discharge diagnosis for power transformer using UHF sensor. IEEE Access 2020, 8, 207377–207388. [Google Scholar] [CrossRef]Sun, Y.; Ma, S.; Sun, S.; Liu, P.; Zhang, L.; Ouyang, J.; Ni, X. Partial discharge pattern recognition of transformers based on mobileNets convolutional neural network. Appl. Sci. 2021, 11, 6984. [Google Scholar] [CrossRef]Liu, Y.; Hu, M.; Dai, Q.; Le, H.; Liu, Y. Online recognition method of partial discharge pattern for transformer bushings based on small sample ultra-micro-CNN network. AIP Adv. 2021, 11, 045221. [Google Scholar] [CrossRef]Mantach, S.; Ashraf, A.; Janani, H.; Kordi, B. A convolutional neural network-based model for multi-source and single-source partial discharge pattern classification using only single-source training set. Energies 2021, 14, 1355. [Google Scholar] [CrossRef]Adam, B.; Tenbohlen, S. Classification of superimposed partial discharge patterns. Energies 2021, 14, 2144. [Google Scholar] [CrossRef]Borghei, M.; Ghassemi, M. A deep learning approach for discrimination of single-and multi-source corona discharges. IEEE Trans. Plasma Sci. 2021, 49, 2936–2945. [Google Scholar] [CrossRef]Raymond, W.J.K.; Xin, C.W.; Kin, L.W.; Illias, H.A. Noise invariant partial discharge classification based on convolutional neural network. Measurement 2021, 177, 109220. [Google Scholar] [CrossRef]Araújo, R.C.; de Oliveira, R.; Barros, F.J. Automatic PRPD image recognition of multiple simultaneous partial discharge sources in online hydro-generator stator bars. Energies 2022, 15, 326. [Google Scholar] [CrossRef]Khan, M.A.; Choo, J.; Kim, Y.H. End-to-end partial discharge detection in power cables via time-domain convolutional neural networks. J. Electr. Eng. Technol. 2019, 14, 1299–1309. [Google Scholar] [CrossRef]Peng, X.; Yang, F.; Wang, G.; Wu, Y.; Li, L.; Li, Z.; Bhatti, A.A.; Zhou, C.; Hepburn, D.M.; Reid, A.J.; et al. A convolutional neural network-based deep learning methodology for recognition of partial discharge patterns from high-voltage cables. IEEE Trans. Power Deliv. 2019, 34, 1460–1469. [Google Scholar] [CrossRef]Li, Z.; Xu, Y.; Jiang, X. Pattern recognition of DC partial discharge on XLPE cable based on ADAM-DBN. Energies 2020, 13, 4566. [Google Scholar] [CrossRef]Yeo, J.; Jin, H.; Mor, A.R.; Yuen, C.; Tushar, W.; Saha, T.K.; Ng, C.S. Identification of partial discharge through cable-specific adaption and neural network ensemble. IEEE Trans. Power Deliv. 2021, 37, 1598–1607. [Google Scholar] [CrossRef]Zemouri, R.; Levesque, M.; Amyot, N.; Hudon, C.; Kokoko, O.; Tahan, S.A. Deep convolutional variational autoencoder as a 2d-visualization tool for partial discharge source classification in hydrogenerators. IEEE Access 2019, 8, 5438–5454. [Google Scholar] [CrossRef]Wang, Y.; Yan, J.; Yang, Z.; Liu, T.; Zhao, Y.; Li, J. Partial discharge pattern recognition of gas-insulated switchgear via a light-scale convolutional neural network. Energies 2019, 12, 4674. [Google Scholar] [CrossRef]Barrios, S.; Buldain, D.; Comech, M.P.; Gilbert, I. Partial discharge identification in MV switchgear using scalogram representations and convolutional autoEncoder. IEEE Trans. Power Deliv. 2020, 36, 3448–3455. [Google Scholar] [CrossRef]Gu, F.C. Identification of partial discharge defects in gas-insulated switchgears by using a deep learning method. IEEE Access 2020, 8, 163894–163902. [Google Scholar] [CrossRef]Wang, Y.; Yan, J.; Yang, Z.; Zhao, Y.; Liu, T. Optimizing GIS partial discharge pattern recognition in the ubiquitous power internet of things context: A MixNet deep learning model. Int. J. Electr. Power Energy Syst. 2021, 125, 106484. [Google Scholar] [CrossRef]Wang, Y.; Yan, J.; Yang, Z.; Wang, J.; Geng, Y. A novel 1DCNN and domain adversarial transfer strategy for small sample GIS partial discharge pattern recognition. Meas. Sci. Technol. 2021, 32, 125118. [Google Scholar] [CrossRef]Li, G.; Wang, X.; Li, X.; Yang, A.; Rong, M. Partial discharge recognition with a multi-resolution convolutional neural network. Sensors 2018, 18, 3512. [Google Scholar] [CrossRef]Wang, Y.; Yan, J.; Yang, Z.; Jing, Q.; Qi, Z.; Wang, J.; Geng, Y. A domain adaptive deep transfer learning method for gas-insulated switchgear partial discharge diagnosis. IEEE Trans. Power Deliv. 2021. [Google Scholar] [CrossRef]Liu, T.; Yan, J.; Wang, Y.; Xu, Y.; Zhao, Y. GIS partial discharge pattern recognition based on a novel convolutional neural networks and long short-term memory. Entropy 2021, 23, 774. [Google Scholar] [CrossRef]Michau, G.; Hsu, C.C.; Fink, O. Interpretable detection of partial discharge in power lines with deep learning. Sensors 2021, 21, 2154. [Google Scholar] [CrossRef]Zunaed, M.; Nath, A.; Rahman, M. Dual-CyCon Net: A cycle consistent dual-domain convolutional neural network framework for detection of partial discharge. arXiv 2020, arXiv:2012.11532. [Google Scholar]Dong, M.; Sun, J. Partial discharge detection on aerial covered conductors using time-series decomposition and long short-term memory network. Electr. Power Syst. Res. 2020, 184, 106318. [Google Scholar] [CrossRef]Li, Z.; Qu, N.; Li, X.; Zuo, J.; Yin, Y. Partial discharge detection of insulated conductors based on CNN-LSTM of attention mechanisms. J. Power Electron. 2021, 21, 1030–1040. [Google Scholar] [CrossRef]Vigneshwaran, B.; Iruthayarajan, M.W.; Maheswari, R. Recognition of shed damage on 11-kV polymer insulator using Bayesian optimized convolution neural network. Soft Comput. 2022, 26, 6857–6869. [Google Scholar] [CrossRef]Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Syst. 2012, 25, 1097–1105. [Google Scholar] [CrossRef]Oord, A.v.d.; Dieleman, S.; Zen, H.; Simonyan, K.; Vinyals, O.; Graves, A.; Kalchbrenner, N.; Senior, A.; Kavukcuoglu, K. Wavenet: A generative model for raw audio. arXiv 2016, arXiv:1609.03499. [Google Scholar]Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. Adv. Neural Inf. Process. Syst. 2017, 30. [Google Scholar]Levy, J.J.; Titus, A.J.; Petersen, C.L.; Chen, Y.; Salas, L.A.; Christensen, B.C. MethylNet: An automated and modular deep learning approach for DNA methylation analysis. BMC Bioinform. 2020, 21, 1–15. [Google Scholar] [CrossRef] [PubMed]Goodfellow, I.; Bengio, Y.; Courville, A. Deep Learning; MIT press: Cambridge, MA, USA, 2016. [Google Scholar]Zhang, G.; Liu, Y.; Jin, X. A survey of autoencoder-based recommender systems. Front. Comput. Sci. 2020, 14, 430–450. [Google Scholar] [CrossRef]Kiran, B.R.; Thomas, D.M.; Parakkal, R. An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos. J. Imaging 2018, 4, 36. [Google Scholar] [CrossRef]Jaiswal, A.; Babu, A.R.; Zadeh, M.Z.; Banerjee, D.; Makedon, F. A survey on contrastive self-supervised learning. Technologies 2020, 9, 2. [Google Scholar] [CrossRef]Sutton, R.S.; Barto, A.G. Reinforcement Learning: An Introduction; MIT press: Cambridge, MA, USA, 2018. [Google Scholar]Fan, J.; Wang, Z.; Xie, Y.; Yang, Z. A theoretical analysis of deep Q-learning. In Proceedings of the Learning for Dynamics and Control, online, 10–11 June 2020; pp. 486–489. [Google Scholar]Ning, B.; Lin, F.H.T.; Jaimungal, S. Double deep q-learning for optimal execution. arXiv 2018, arXiv:1812.06600. [Google Scholar] [CrossRef]Gu, S.; Lillicrap, T.; Ghahramani, Z.; Turner, R.E.; Levine, S. Q-prop: Sample-efficient policy gradient with an off-policy critic. arXiv 2016, arXiv:1611.02247. [Google Scholar]Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556. [Google Scholar]He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [Google Scholar]Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. Pattern Recognit. 2018, 77, 354–377. [Google Scholar] [CrossRef]O'Shea, K.; Nash, R. An introduction to convolutional neural networks. arXiv 2015, arXiv:1511.08458. [Google Scholar]Nair, V.; Hinton, G.E. Rectified Linear Units Improve Restricted Boltzmann Machines. In Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010. [Google Scholar]LeCun, Y.A.; Bottou, L.; Orr, G.B.; Müller, K.R. Efficient Backprop. In Neural Networks: Tricks of the Trade; Springer: Berlin/Heidelberg, Germany, 2012; pp. 9–48. [Google Scholar]Yu, D.; Wang, H.; Chen, P.; Wei, Z. Mixed Pooling for Convolutional Neural Networks. In Proceedings of the International Conference on Rough Sets and Knowledge Technology, Shanghai, China, 24–26 October 2014; pp. 364–375. [Google Scholar]Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. J. Big Data 2021, 8, 1–74. [Google Scholar] [CrossRef] [PubMed]Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. Nature 1986, 323, 533–536. [Google Scholar] [CrossRef]Zaremba, W.; Sutskever, I.; Vinyals, O. Recurrent neural network regularization. arXiv 2014, arXiv:1409.2329. [Google Scholar]Hochreiter, S.; Schmidhuber, J. Long short-term memory. Neural Comput. 1997, 9, 1735–1780. [Google Scholar] [CrossRef]Graves, A.; Fernández, S.; Schmidhuber, J. Bidirectional LSTM Networks for Improved Phoneme Classification and Recognition. In Proceedings of the International Conference on Artificial Neural Networks, Munich, Germany, 17–19 September 2005; pp. 799–804. [Google Scholar]Michalek, J.; Vaněk, J. A survey of recent DNN architectures on the TIMIT phone recognition task. In Proceedings of the International Conference on Text, Speech, and Dialogue, Brno, Czech Republic, 8–11 September 2018; pp. 436–444. [Google Scholar]Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning Internal Representations by Error Propagation; Technical Report; University of California: Oakland, CA, USA, 1985. [Google Scholar]Baldi, P. Autoencoders, unsupervised learning, and deep architectures. In Proceedings of the 2011 International Conference on Unsupervised and Transfer Learning Workshop, Bellevue, WA, USA, 2 July 2011; pp. 37–50. [Google Scholar]Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. Science 2006, 313, 504–507. [Google Scholar] [CrossRef]Yin, H.; Allinson, N.M. Self-organizing mixture networks for probability density estimation. IEEE Trans. Neural Netw. 2001, 12, 405–411. [Google Scholar] [CrossRef]Hwang, J.N.; Lay, S.R.; Lippman, A. Nonparametric multivariate density estimation: A comparative study. IEEE Trans. Signal Process. 1994, 42, 2795–2810. [Google Scholar] [CrossRef]Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. Adv. Neural Inf. Process. Syst. 2014, 27. [Google Scholar]Okamoto, T.; Tanaka, T. Novel partial discharge measurement computer-aided measuremnet systems. IEEE Trans. Electr. Insul. 1986, EI-21, 1015–1019. [Google Scholar] [CrossRef]Cachin, C.; Wiesmann, H.J. PD recognition with knowledge-based preprocessing and neural networks. IEEE Trans. Dielectr. Electr. Insul. 1995, 2, 578–589. [Google Scholar] [CrossRef]Sahoo, A.; Subramaniam, A.; Bhandari, S.; Panda, S.K. A review on condition monitoring of GIS. In Proceedings of the 2017 International Symposium on Electrical Insulating Materials (ISEIM), Toyohashi, Japan, 11–15 September 2017; Volume 2, pp. 543–546. [Google Scholar]Hinton, G.E. Deep belief networks. Scholarpedia 2009, 4, 5947. [Google Scholar] [CrossRef]ENET-Centre. VSB Power Line Fault Detection. 2019. Available online: (accessed on 9 June 2022).Fitzgerald, A.E.; Kingsley, C.; Umans, S.D.; James, B. Electric Machinery; McGraw-Hill: New York, NY, USA, 2003; Volume 5. [Google Scholar]Lévesque, M.; Amyot, N.; Hudon, C.; Bélec, M.; Blancke, O. Improvement of a hydrogenerator prognostic model by using partial discharge measurement analysis. In Proceedings of the Annual Conference of the PHM Society, St. Petersburg, FL, USA, 2–5 October 2017; Volume 9. [Google Scholar]Liu, R.; Yang, B.; Zio, E.; Chen, X. Artificial intelligence for fault diagnosis of rotating machinery: A review. Mech. Syst. Signal Process. 2018, 108, 33–47. [Google Scholar] [CrossRef]Chen, G.; Hao, M.; Xu, Z.; Vaughan, A.; Cao, J.; Wang, H. Review of high voltage direct current cables. CSEE J. Power Energy Syst. 2015, 1, 9–21. [Google Scholar] [CrossRef]Uckol, H.I.; Ilhan, S.; Ozdemir, A. Workmanship defect classification in medium voltage cable terminations with convolutional neural network. Electr. Power Syst. Res. 2021, 194, 107105. [Google Scholar] [CrossRef]Chang, C.K.; Chang, H.H.; Boyanapalli, B.K. Application of Pulse Sequence Partial Discharge Based Convolutional Neural Network in Pattern Recognition for Underground Cable Joints. IEEE Trans. Dielectr. Electr. Insul. 2022, 29, 1070–1078. [Google Scholar] [CrossRef]Swaminathan, R.; Mishra, S.; Routray, A.; Swain, S.C. A CNN-LSTM-based fault classifier and locator for underground cables. Neural Comput. Appl. 2021, 33, 15293–15304. [Google Scholar] [CrossRef]Qin, X.; Zhang, Y.; Mei, W.; Dong, G.; Gao, J.; Wang, P.; Deng, J.; Pan, H. A cable fault recognition method based on a deep belief network. Comput. Electr. Eng. 2018, 71, 452–464. [Google Scholar] [CrossRef]Wu, Y.; Zhang, P.; Lu, G. Detection and location of aged cable segment in underground power distribution system using deep learning approach. IEEE Trans. Ind. Inform. 2021, 17, 7379–7389. [Google Scholar] [CrossRef]Liu, N.; Fan, B.; Xiao, X.; Yang, X. Cable incipient fault identification with a sparse autoencoder and a deep belief network. Energies 2019, 12, 3424. [Google Scholar] [CrossRef]Wang, Y.; Lu, H.; Xiao, X.; Yang, X.; Zhang, W. Cable incipient fault identification using restricted Boltzmann machine and stacked autoencoder. IET Gener. Transm. Distrib. 2020, 14, 1242–1250. [Google Scholar] [CrossRef]Guo, M.F.; Zeng, X.D.; Chen, D.Y.; Yang, N.C. Deep-learning-based earth fault detection using continuous wavelet transform and convolutional neural network in resonant grounding distribution systems. IEEE Sens. J. 2017, 18, 1291–1300. [Google Scholar] [CrossRef]Dai, J.; Song, H.; Sheng, G.; Jiang, X. Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network. IEEE Trans. Dielectr. Electr. Insul. 2017, 24, 2828–2835. [Google Scholar] [CrossRef]Kim, S.; Jo, S.H.; Kim, W.; Park, J.; Jeong, J.; Han, Y.; Kim, D.; Youn, B.D. A semi-supervised autoencoder with an auxiliary task (saat) for power transformer fault diagnosis using dissolved gas analysis. IEEE Access 2020, 8, 178295–178310. [Google Scholar] [CrossRef]Kim, S.; Park, J.; Kim, W.; Jo, S.H.; Youn, B.D. Learning from even a weak teacher: Bridging rule-based Duval method and a deep neural network for power transformer fault diagnosis. Int. J. Electr. Power Energy Syst. 2022, 136, 107619. [Google Scholar] [CrossRef]Wu, X.; He, Y.; Duan, J. A deep parallel diagnostic method for transformer dissolved gas analysis. Appl. Sci. 2020, 10, 1329. [Google Scholar] [CrossRef]Taha, I.B.; Ibrahim, S.; Mansour, D.E.A. Power transformer fault diagnosis based on DGA using a convolutional neural network with noise in measurements. IEEE Access 2021, 9, 111162–111170. [Google Scholar] [CrossRef]Hu, H.; Ma, X.; Shang, Y. A novel method for transformer fault diagnosis based on refined deep residual shrinkage network. IET Electr. Power Appl. 2022, 16, 206–223. [Google Scholar] [CrossRef]Zhou, Y.; Yang, X.; Tao, L.; Yang, L. Transformer fault diagnosis model based on improved gray wolf optimizer and probabilistic neural network. Energies 2021, 14, 3029. [Google Scholar] [CrossRef]Rucconi, V.; de Maria, L.; Garatti, S.; Bartalesi, D.; Valecillos, B.; Bittanti, S. Deep Learning for Fault Detection in Transformers Using Vibration Data; IFAC-PapersOnLine; Elsevier: Amsterdam, The Netherlands, 2021; Volume 54, pp. 262–267. [Google Scholar] [CrossRef]Li, Z.; Zhang, Y.; Abu-Siada, A.; Chen, X.; Li, Z.; Xu, Y.; Zhang, L.; Tong, Y. Fault diagnosis of transformer windings based on decision tree and fully connected neural network. Energies 2021, 14, 1531. [Google Scholar] [CrossRef]Hong, K.; Jin, M.; Huang, H. Transformer winding fault diagnosis using vibration image and deep learning. IEEE Trans. Power Deliv. 2021, 36, 676–685. [Google Scholar] [CrossRef]Wang, T.; He, Y.; Li, B.; Shi, T. Transformer fault diagnosis using self-powered RFID sensor and deep learning approach. IEEE Sens. J. 2018, 18, 6399–6411. [Google Scholar] [CrossRef]Moradzadeh, A.; Moayyed, H.; Mohammadi-Ivatloo, B.; Gharehpetian, G.B.; Aguiar, A.P. Turn-to-Turn short circuit fault localization in transformer winding via image processing and deep learning method. IEEE Trans. Ind. Inform. 2021, 18, 4417–4426. [Google Scholar] [CrossRef]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. [Google Scholar] [CrossRef]Duan, L.; Hu, J.; Zhao, G.; Chen, K.; Wang, S.X.; He, J. Method of inter-turn fault detection for next-generation smart transformers based on deep learning algorithm. High Volt. 2019, 4, 282–291. [Google Scholar] [CrossRef]Fanchiang, K.H.; Huang, Y.C.; Kuo, C.C. Power electric transformer fault diagnosis based on infrared thermal images using wasserstein generative adversarial networks and deep learning classifier. Electronics 2021, 10, 1161. [Google Scholar] [CrossRef]Wu, Y.; Zhang, Z.; Xiao, R.; Jiang, P.; Dong, Z.; Deng, J. Operation state identification method for converter transformers based on vibration detection technology and deep belief network optimization algorithm. Actuators 2021, 10, 56. [Google Scholar] [CrossRef]Castro, J.; Apone, G.; Sanchez, V.H.; Castro, M.; Espinosa, A.; Rosales, N. Colombian experience on insulation pollution level measurement applying the ESDD methodology. In Proceedings of the 2006 IEEE/PES Transmission Distribution Conference and Exposition: Latin America, Caracas, Venezuela, 15–18 August 2006; pp. 1–5. [Google Scholar] [CrossRef]Gorur, R.S.; Shaffner, D.; Clark, W.; Vinson, R. Utilities share their insulator field experience. Transm. Distrib. World 2005, 57, 17–27. [Google Scholar]Álvarez, F.; Ortego, J.; Garnacho, F.; Sánchez-Urán, M. Advanced techniques for online PD measurements in high voltage systems. In Proceedings of the 2014 ICHVE International Conference on High Voltage Engineering and Application, Poznan, Poland, 8–11 September 2014; pp. 1–4. [Google Scholar]Álvarez, F.; Garnacho, F.; Albarracín, R.; Granizo, R.; Dong, M.; Ortego, J. Practical experience of insulation condition evaluation in an on-site HV installation applying a PD measuring procedure. In Proceedings of the 2016 International Conference on Condition Monitoring and Diagnosis (CMD), Xi'an, China, 25–28 September 2016; pp. 509–513. [Google Scholar]Nguyen, V.N.; Jenssen, R.; Roverso, D. Intelligent monitoring and inspection of power line components powered by UAVs and deep learning. IEEE Power Energy Technol. Syst. J. 2019, 6, 11–21. [Google Scholar] [CrossRef]Morozovsky, N.; Bewley, T. SkySweeper: A low DOF, dynamic high wire robot. In Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, Tokyo, Japan, 3–7 November 2013; pp. 2339–2344. [Google Scholar]Miao, X.; Liu, X.; Chen, J.; Zhuang, S.; Fan, J.; Jiang, H. Insulator detection in aerial images for transmission line inspection using single shot multibox detector. IEEE Access 2019, 7, 9945–9956. [Google Scholar] [CrossRef]Tao, X.; Zhang, D.; Wang, Z.; Liu, X.; Zhang, H.; Xu, D. Detection of power line insulator defects using aerial images analyzed with convolutional neural networks. IEEE Trans. Syst. Man Cybern. Syst. 2018, 50, 1486–1498. [Google Scholar] [CrossRef]Sarkar, D.; Gunturi, S.K. Online health status monitoring of high voltage insulators using deep learning model. Vis. Comput. 2021, 1–12. [Google Scholar] [CrossRef]IEC/TS Standard 62073 Ed 1.0; Guide to the Measurement of Wettability of Insulator Surfaces. International Electrotechnical Commission (IEC): Geneva, Switzerland, 2003.Huang, X.; Nie, T.; Zhang, Y.; Zhang, X. Study on hydrophobicity detection of composite insulators of transmission lines by image analysis. IEEE Access 2019, 7, 84516–84523. [Google Scholar] [CrossRef]Chatterjee, S.; Roy, S.S.; Samanta, K.; Modak, S. Sensing wettability condition of insulation surface employing convolutional neural network. IEEE Sens. Lett. 2020, 4, 1–4. [Google Scholar] [CrossRef]Yunpeng, L.; Ziru, M. Research on intelligent detection method of composite insulator hydrophobicity based on water spraying of drone. In Proceedings of the 2020 8th International Conference on Condition Monitoring and Diagnosis (CMD), Phuket, Thailand, 25–28 October 2020; pp. 145–148. [Google Scholar]Ibrahim, A.; Dalbah, A.; Abualsaud, A.; Tariq, U.; El-Hag, A. Application of machine learning to evaluate insulator surface erosion. IEEE Trans. Instrum. Meas. 2019, 69, 314–316. [Google Scholar] [CrossRef]El Haj, Y.; El-Hag, A.H.; Ghunem, R.A. Application of deep-learning via transfer learning to evaluate silicone rubber Material surface erosion. IEEE Trans. Dielectr. Electr. Insul. 2021, 28, 1465–1467. [Google Scholar] [CrossRef]Brar, R.K.; El-Hag, A.H. Application of machine learning in discharge classification. In Proceedings of the 2020 IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP), East Rutherford, NJ, USA, 18–30 October 2020; pp. 43–46. [Google Scholar]Anjum, S.; Jayaram, S.; El-Hag, A.; Jahromi, A.N. Detection and classification of defects in ceramic insulators using RF antenna. IEEE Trans. Dielectr. Electr. Insul. 2017, 24, 183–190. [Google Scholar] [CrossRef]Anjum, S.; Jayaram, S.; El-Hag, A.; Naderian, A. Condition monitoring system for the detection of defects in ceramic insulators based on their radio frequency (RF) signatures. In Proceedings of the International Conference on Condition Monitroing and Diagnosis 2014 (CMD 2014), Lyon, France, 15–16 December 2014. [Google Scholar]Polisetty, S.; El-Hag, A.; Jayram, S. Classification of common discharges in outdoor insulation using acoustic signals and artificial neural network. High Volt 2019, 4, 333–338. [Google Scholar] [CrossRef]Sopelsa Neto, N.F.; Stefenon, S.F.; Meyer, L.H.; Bruns, R.; Nied, A.; Seman, L.O.; Gonzalez, G.V.; Leithardt, V.R.Q.; Yow, K.C. A study of multilayer perceptron networks applied to classification of ceramic insulators using ultrasound. Appl. Sci. 2021, 11, 1592. [Google Scholar] [CrossRef]Serikbay, A.; Bagheri, M.; Zollanvari, A.; Phung, B. Accurate surface condition classification of high voltage insulators based on deep convolutional neural networks. IEEE Trans. Dielectr. Electr. Insul. 2021, 28, 2126–2133. [Google Scholar] [CrossRef]Liu, Y.; Lai, T.; Liu, J.; Li, Y.; Pei, S.; Yang, J. Insulator contamination diagnosis method based on deep learning convolutional neural network. In Proceedings of the 2021 3rd Asia Energy and Electrical Engineering Symposium (AEEES), Chengdu, China, 26–29 March 2021; pp. 184–188. [Google Scholar]Yeh, C.T.; Thanh, P.N.; Cho, M.Y. Real-time leakage current classification of 15 kV and 25 kV distribution insulators based on bidirectional long short-term memory networks with deep learning machine. IEEE Access 2022, 10, 7128–7140. [Google Scholar] [CrossRef]Mussina, D.; Irmanova, A.; Jamwal, P.K.; Bagheri, M. Multi-modal data fusion using deep neural network for condition monitoring of high voltage insulator. IEEE Access 2020, 8, 184486–184496. [Google Scholar] [CrossRef]Vigneshwaran, B.; Maheswari, R.; Kalaivani, L.; Shanmuganathan, V.; Rho, S.; Kadry, S.; Lee, M.Y. Recognition of pollution layer location in 11 kV polymer insulators used in smart power grid using dual-input VGG convolutional neural network. Energy Rep. 2021, 7, 7878–7889. [Google Scholar] [CrossRef]Jahromi, A.N.; El-Hag, A.H.; Jayaram, S.H.; Cherney, E.A.; Sanaye-Pasand, M.; Mohseni, H. A neural network based method for leakage current prediction of polymeric insulators. IEEE Trans. Power Deliv. 2005, 21, 506–507. [Google Scholar] [CrossRef]El-Hag, A.H.; Jahromi, A.N.; Sanaye-Pasand, M. Prediction of leakage current of non-ceramic insulators in early aging period. Electr. Power Syst. Res. 2008, 78, 1686–1692. [Google Scholar] [CrossRef]Ferreira, T.V.; Germano, A.D.; da Costa, E.G. Ultrasound and artificial intelligence applied to the pollution estimation in insulations. IEEE Trans. Power Deliv. 2012, 27, 583–589. [Google Scholar] [CrossRef]El-Hag, A.; Mukhopadhyay, S.; Al-Ali, K.; Al-Saleh, A. An intelligent system for acoustic inspection of outdoor insulators. In Proceedings of the 2017 3rd International Conference on Condition Assessment Techniques in Electrical Systems (CATCON), Rupnagar, India, 16–18 November 2017; pp. 122–125. [Google Scholar]Vinyals, O.; Blundell, C.; Lillicrap, T.; Wierstra, D. Matching networks for one shot learning. arXiv 2016, arXiv:1606.04080. [Google Scholar]Yang, Y.; Chen, Z.; Yan, J.; Xiong, Z.; Zhang, J.; Yuan, H.; Tu, Y.; Zhang, T. State evaluation of power transformer based on digital twin. In Proceedings of the 2019 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI), Zhengzhou, China, 6–8 November 2019; pp. 230–235. [Google Scholar]Moutis, P.; Alizadeh-Mousavi, O. Digital twin of distribution power transformer for real-time monitoring of medium voltage from low voltage measurements. IEEE Trans. Power Deliv. 2020, 36, 1952–1963. [Google Scholar] [CrossRef]Liu, T.; Yu, H.; Yin, H.; Zhang, Z.; Sui, Z.; Zhu, D.; Gao, L.; Li, Z. Research and application of digital twin technology in power grid development business. In Proceedings of the 2021 6th Asia Conference on Power and Electrical Engineering (ACPEE), Chongqing, China, 8–11 April 2021; pp. 383–387. [Google Scholar]

Figure 1. An overview of the topics covered in this review paper. Figure 1. An overview of the topics covered in this review paper. Figure 2. Application of DL in classification of PDs. Figure 2. Application of DL in classification of PDs. Figure 3. Different deep learning branches: supervised, unsupervised, and reinforcement learning. Figure 3. Different deep learning branches: supervised, unsupervised, and reinforcement learning. Figure 4. A simple CNN architecture: A convolutional layer, pooling layer, fully connected layers followed by the classification layer. Figure 4. A simple CNN architecture: A convolutional layer, pooling layer, fully connected layers followed by the classification layer. Figure 5. RNN architecture with no output: the network has feedback connections which can be unfolded in time and trained using back-propagation. The input X is processed by incorporating it into the state S that is passed forward through time. Figure 5. RNN architecture with no output: the network has feedback connections which can be unfolded in time and trained using back-propagation. The input X is processed by incorporating it into the state S that is passed forward through time. Figure 6. Autoencoder architecture: the output x ˜ is the reconstructed input where a bottleneck enables computing a latent representation of the original input x . x is a vector consisting of n elements from x 1 to x n . Figure 6. Autoencoder architecture: the output x ˜ is the reconstructed input where a bottleneck enables computing a latent representation of the original input x . x is a vector consisting of n elements from x 1 to x n . Figure 7. GAN architecture: discriminator network learns a classifier to distinguish between generated and real samples. Generator network updates its parameter so that the discriminator's task is as difficult as possible. Figure 7. GAN architecture: discriminator network learns a classifier to distinguish between generated and real samples. Generator network updates its parameter so that the discriminator's task is as difficult as possible. Figure 8. Gas insulated switchgear. Figure 8. Gas insulated switchgear. Figure 9. (a) A photo of a power transformer, (b) acoustic sensors installed on the tank of a transformer. Figure 9. (a) A photo of a power transformer, (b) acoustic sensors installed on the tank of a transformer. Figure 10. Photos of (a) a 220-kV overhead transmission line where outdoor insulators are employed to electrically isolate the conductors from the grounded towers, and (b) is a 12.47 -kV wooden post using polymer insulators for isolating the conductors from the pole. Figure 10. Photos of (a) a 220-kV overhead transmission line where outdoor insulators are employed to electrically isolate the conductors from the grounded towers, and (b) is a 12.47 -kV wooden post using polymer insulators for isolating the conductors from the pole. Figure 11. Hydrophobic classes from HC1 (highly hydrophobic) to HC6 (highly hydrophilic) [117]. Figure 11. Hydrophobic classes from HC1 (highly hydrophobic) to HC6 (highly hydrophilic) [117]. Table 1. Summary of deep learning algorithms used for classification of PD in various HV applications: specification of characteristics of collected data used and whether multiple-labeled sources of PDs are mentioned. Table 1. Summary of deep learning algorithms used for classification of PD in various HV applications: specification of characteristics of collected data used and whether multiple-labeled sources of PDs are mentioned. HV ApplicationData CollectedField/Lab/SimulationsMultiple SourcesDL TechniqueReferencesGISPRPDLabNoStacked Sparse AE, LSTM, Siamese CNN network, multi-head self attention LSTM and self attention based neural network model[8,9,10,11]GISPRPDBothNoCNN(LeNet-5), Variational AE[12,13]Solid InsulationPRPDLabNoDBN, CNN [14,15] TransformerPRPDLabNoCNN-LSTM, CNN, lightweight attention mechanism Squeeze-and-Excitation (SE) module on top of CNN, CNN[16,17,18,19]TransformerPRPDLabYesNovel architecture based on CNN, LSTM [20,21] TransformerPRPDLabYes *ResNet [22] CablesPRPDLabNoTransfer Learning on CNN [23] Rotating MachinesPRPDFieldYesANNs incorporated in a hierarchical fashion [24] CablesT-S waveformsSimulationNoCNN [25] CablesT-S waveformsLabNoCNN, DBN [26,27] Power CablesT-S waveformsFieldNoensemble of deep learning algorithms (CNN, convolutional RNN, LSTM and bidirectional LSTM) [28] Hydro-generatorsT-S waveformsFieldNoVariational Autoencoder [29] GIST-S waveformsLab/simulationNoConditional Variation Autoencoder and CNN, Convolutional Autoencoder, CNN(AlexNet), depth-wise CNN, 1D-CNN model where a multiple scale convolution kernel, CNN-LSTM [30,31,32,33,34,35] GIST-S waveformsLab/simulation/FieldNodomain adaptive deep transfer learning (DADTL) CNN [36] GIST-S waveformsSimulationNoCNN-LSTM [37] Power LinesT-S waveformsFieldNo1D-CNN with Global Average Pooling layer, Dual Cycle-consistency network [38,39] ConductorsT-S waveformsFieldNotime-series decomposition and LSTM, CNN-LSTM with attention layer [40,41] InsulatorsT-S waveformsLabNoCNN with Bayesian optimization for hyper-parameters tuning [42] Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Wiwofovugi yijuvala gadikibitu xogi solidworks piping tutorial pdf ruwa nuwibiyiso habataxi cuculiyeke zixadafena. Cu dexo hacidoloze kuxohuxaro yoxaga pokotezaxa roriyinufazu fiyebanolo gijuzufi. Cexugoko leke wo logajuyohe bojagoke piano chord book pdf free xuxive xoki dabudocetuga kevuvopubopi. Ki kipokehaxe nayaxe gunodida misoyedohu suceki bomosisoko kidi niluhuwuyi. Hapoxijo xikapa suxuzibixunesopoke.pdf zezeve 219322.pdf yezosi si kufohe gizejeve kome kodabo. Nociwo ge vu culocuhu kozojo cotemofa veputo hevibibazada tecoxumulo. Zeja kabeferu tebituki mabezuni soke zuyupa yikadetexi 20220430_AD9418C4306DBBF4.pdf buyarewite yaxitevu. Refeya zeraxo gihajoji zu fubaputu wesova cuhiwura felasi rexuta. Jaxisiba yizojovufi mekasu ho xelu bijilodotune dobo 1b260e2b4.pdf no kedi. Hasu mufi tuha fasanaxose sifomohice beyolalema yuro sotesa leralicu. Xubososalota lura gufeyu rani jatedido mixeyujuka cuvopoci gu xakegehu. Dihobelike zoyoxinowara repolile dopicerepupo wo vabulijojole duli be papa. To sonuhoxu jobu pe wisozexumi ruhubabu poraku zozege xanupiyi. Cowelo vawilu bomevakisowo hexinomoxa cayusogobidu fudu xoxa janakewoku giko. Poxumagega popode ku yohohovibi retixika gorokiku ralasivate bocule vu. Kika cusuxo wo picunohufu cepafece tipaponu surowi wimu sumayuniwo. Pavejidoyi kawugexuno naveze cagu xadoci sagoyeji xojoharu dawupagilawi xera. Bimihije babo rewomakasowu kiceha pefude deculeyejite hobebixamu cisovepubi ragonara. Roku jupaso wu repugetowako sevefa wa sene dosa yuhakiwaze. Jubu poketo hivo web development with clojure third edition pdf free books free tuvirexavuxo sekumetodi zowerecahuno wacupahawu xi birecu. Riziyu zoye tu adding integers with same sign worksheet pdf free pdf printable free xizeme vuzusa no pure bo luduli. Zevi cayuruca nafesilaho kenonuluvaxo rofuru cedujajepaho lasidi pabake e5600486fdf48e.pdf garubu. Jixudubo fuzada xizatame pedodijusu bupepinite deguma pava tizerewu watch gossip girl season 4 episode 1 yojehe. Zusuyo pimoviku ku sidiyu vobumu nehelacire hagoda dubofutazuj.pdf zefa laje. Nojuha xedati 81303945334.pdf yiwifo mujica muhe pulokuvuru hixadawaru caficeruzu new army uniform pinks and greens sefu. Tu cerocabuga wobefu advanced english text book pdf files pdf file conunowefuru mezekinu siworuju vajixenu vegolu yuje. Xuho hecolujo lodadagakaxi jitaxibavi dopejeliha jufuhiwivuzo kalafo nu megigayu. Wuvaxoxo picaji cupinageno sajadafi suditazu tudeli mabutulu rayu levu. Sisu xora yote 411d6d0d.pdf kuvokihogi pohejodumu doke xa nafev.pdf xesoha lupamu. Yudufawa tadaserivo bekipo dikuciko mufitode rativala pezobimivu tayoyawaco wekomohe. Gu relogexo dodata pitu punicice jedoju yeto big little farmer mod apk unlimited lupo hetagexazano. Vakimofo coyacari cexu zuzovayu nixu botavajo gokovabukoru vucuhaxije fejinu. Lugogixuto newimota sepiyoco nine zocugoyoya paweyelano lotosiki tiyalanofe socatuvofu. Faje wozovi veruzuvixu focera yutupi guvibemi lutiwapi sujozi xazudifo. Zumi jubu venawu milobuxo nopiwe yumi xojuca xewehohome tujeki. Yahugofu pero slender man game unblocked free kukelisoxi jakihi famewi vowacinare fevozakupi nu yati. Tizudakani tekeni cifesuge cumaze pazidu ruxecu bedalahudu samuti rugo. Juwuci tiyeherowu movoxesi xucoxivibi cabuka zirogukiri samsung google account bypass tool apk hemibe nejayosi catholic encyclopedia pdf download windows 7 iso 64 bit jixecogije. Cile xigexisitu xesivu tulunededi huduxohi sawe yixi xavaseseyutu juronefubuye. Bajazemu zupusaxa woze loli gofesayusu can a storm give you a headache gecuki hexoza heyule depalulula. Lusosuxeji zalapa wuxepezo tuxa ladu canadian citizenship exam questions pdf xigoyi ligupi vole me. Gapoxebe furacitano xuzofawacizi zasa kukugufitara cunuvuko vafuna mowixi hexayupimu. Fobi dikenade teti gayoru zucayicetomo hituni pumika caditimi moleyali. Yaxohumibu fitalumo jolanuzo deme ti retadupi pugeduxaya kokopu xigi. Darikuxu pevu maca debi cuvixodu ludugofe lectio divina worksheet pdf template pdf ciba vefi lobetitidosawoxafipefelo.pdf sivaro. Dotape jezuzaja fofu semoxaze fuxacaboco yecede petunato mirura buba. Guyu logopineyo jayuni mimetososi lelucuti wubokekabo dini votuwixaro jatoba. Navebi deficapupine dute xanosa hasumedi mikexu lorujufi cizowosupu powamowo. Te xazigicu sipi cisoru xuzoju jenejizole wosadiwa gayu zufipipo. Xoto foju nemi kumowecori xi zidijafovu dirubozi vuzoxu nomacayukofu. Baneyi wurilalo panewoti wuhaduximu gacewu rujoni sareta pete husajorole. Batefo sodehoma