Measurement 177 (2021) 109220

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Noise invariant partial discharge classification based on convolutional neural network

Wong Jee Keen Raymond [a,b,*], Chong Wan Xin [a], Lai Weng Kin [a], Hazlee Azil Illias [b]

[a] Department of Electrical and Electronics Engineering, Faculty of Engineering and Technology, Tunku Abdul Rahman University College, 53300 Kuala Lumpur, Malaysia
[b] Department of Electrical Engineering, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia

ARTICLE INFO

Keywords:
Partial discharge
Feature extraction
Convolutional neural network
Pattern recognition
Electrical insulation
Machine learning

ABSTRACT

Partial discharge (PD) pattern recognition is essential since it can help to identify the nature of the insulation defect. Numerous machine learning models have been utilized for PD classification applications in the past. However, traditional machine learning models rely on manual feature extraction to obtain training data. They are usually trained using clean PD data measured in the laboratory but are expected to work on-site where some degree of interference or noise is expected. When tested using clean PD data, most machine learning models can easily achieve above 90% accuracy. However, when tested using PD data overlapped with noise, classification accuracy reduces significantly. In this work, the development of a convolutional neural network (CNN)-based PD classification system using transfer learning was proposed. In order to achieve a more practical performance evaluation, a modified 10-fold cross-validation procedure was used where the CNN-based PD classifier was trained using clean PD data but tested using PD data that has been overlapped by noise. The results showed that CNN-based PD classifier was able to achieve up to 16.90% higher classification accuracy under noise contamination compared to traditional machine learning with manual feature extraction. This shows that the proposed method was able to retain higher classification accuracy in the presence of noise.

1. Introduction

The insulation of high voltage equipment is continuously subjected to electrical stresses that cause performance degradation. In addition to the electrical fields, the insulation also experiences mechanical, thermal, and environmental stresses, which drastically accelerate their ageing process [1]. Most of these degradation factors result in the phenomenon known as PD, tiny ionization processes that occur at the location where the electric field is highly inhomogeneous [2]. If left unchecked, PD will eventually degrade the insulation to the point of electrical breakdown. Hence, PD pattern recognition is beneficial for the diagnosis of insulation defect type [3]. Efficient diagnostic techniques are vital for enhancing the implementation of maintenance activity by minimizing the servicing area [4].

There is a strong correlation between the PD pattern and the nature of the insulation defect. A classification system could be developed using discriminatory features extracted from the PD pattern. From the phase-resolved PD (PRPD) patterns, statistical parameters such as skewness, kurtosis, mean, variance, and cross-correlation can be calculated [5]. There are also some image processing tools such as texture analysis algorithms [6], fractal features [7], wavelet-based image decomposition [8], which can be used to obtain useful features from the PRPD image. Alternatively, by using signal processing tools such as Fourier series analysis [9], Hilbert–Huang transforms [10], or wavelet transform [11], features can be extracted from the time-domain representation. In terms of PD classification systems, traditional machine learning is the most widely used method. Examples of commonly used traditional machine learning-based models include shallow ($\leq$3 layers) artificial neural networks (ANN) [12], support vector machines (SVM) [13], neuro-fuzzy [14], and random forest [15]. A comprehensive review can be found in [16,17], which summarizes the state-of-the-art techniques for PD feature extraction and PD classification methods used by different researchers.

There are two notable research gaps in previous PD classification works. Firstly, many PD classification research done in the past was trained using PD data collected in a laboratory. However, the PD classification system is expected to be used on-site where some variation due to interference or noise is expected. It is shown that when the PD data is overlapped with simulated noise [18] and mixing laboratory data with on-site data [19], the classification accuracy suffers a reduction of up to

* Corresponding author.
E-mail addresses: wongjk@tarc.edu.my (W.J.K. Raymond), laiwk@tarc.edu.my (L.W. Kin), h.illias@um.edu.my (H.A. Illias).

https://doi.org/10.1016/j.measurement.2021.109220
Received 16 November 2020; Received in revised form 24 December 2020; Accepted 22 February 2021
Available online 2 March 2021

20%. PD denoising is an on-going research field that could help in improving PD classification accuracy to a certain extent. However, a perfect and universal denoising standard has yet to be achieved [20]. Therefore, it is crucial to observe the performance of the PD classifier when tested against data overlapped with noise. Secondly, traditional machine learning utilizes algorithms to analyze data, learn from that data to produce educated decisions based on what they have learned. However, manual feature extraction is required for such a process. The extracted feature's correlation with the PD source and its tolerance to variation/noise in input data will directly affect the classification accuracy. Due to the improvement in computer parallel processing hardware, deep learning is gaining popularity. Deep learning algorithms are based on ANN, which has more than two hidden layers that can learn and form sensible predictions or decisions independently, including automatic feature extraction which sets it apart from traditional machine learning.

The trend in PD classification has moved from traditional machine learning towards deep learning in recent years. The earliest use of deep learning for PD classification was done in [21], where a five hidden layer deep neural network (DNN) was used. Long short term memory network (LSTM), which is a kind of recurrent neural network (RNN), was used in [22] to classify PD patterns in gas-insulated switchgear (GIS). Autoencoder was used in [23] to perform automatic feature extraction from PD data. Generative Adversarial Networks (GAN) was used in [24] to generate additional PD training data via data augmentation. A performance comparison was made in [25], where it was found that CNN outperforms DNN and RNN for classifying PD based on ultrasonic signals.

CNN is one of the essential models of deep learning that is widely used for image recognition. Its unique network structure is highly invariant to a certain degree of translation, scaling, and distortion [26]. In addition, it can automatically learn features from a large amount of training data, depict the intrinsic information of data, avoids the process of manual extraction and selection of features. In recent years, more CNN-based methods have been applied in the field of fault diagnosis, where the input data can be represented in image format. A bearing fault diagnosis model based on AlexNet CNN was developed in [27] using time-frequency images of bearing signals. CNN was used in [28] to characterize vibrational signals of wind turbine gearbox to achieve better fault diagnosis.

Since PD can be represented as a 2D PRPD colour image, CNN is used in this work to develop a PD classification system. By using CNN, manual feature extraction can be avoided as CNN will be able to extract features automatically and perform the classification. The CNN's convolutional layers are able to extract intrinsic features from the training images. Hence, CNN extracted features are expected to achieve better noise tolerance compared to traditional feature extraction methods such as statistical features, which makes it more suitable to be used for on-site PD classification. Training a new CNN from scratch requires a massive amount of training data and computing power. Therefore transfer learning is used, where the existing CNN is repurposed to perform a new classification task, which is to recognize PD patterns.

In this work, five PD sources measured from artificial defects created in XLPE cable joints are used. A comparison of five types of CNN architectures and four types of PD image representation is performed to investigate the PD classification system with the least performance degradation in the presence of noise contamination.

2. PD Data

2.1. PD data acquisition and preprocessing

The PD data used in this work were measured from 11 kV cross-linked polyethylene (XLPE) cable joints with artificial defects commonly encountered in the industry using a commercially available PD detector (Omicron MPD600) that is compliant with IEC60270 standard [29]. The IEC60270 standard recommends that the lower frequency (f_1) to be between 30 kHz and 100 kHz, the upper frequency (f_2) to be less than 500 kHz, and the bandwidth ($f_2 - f_1$) to be between 100 kHz and 400 kHz. Thus, the center frequency used for charge integration setting was set to 250 kHz with a bandwidth of 300 kHz ($f_1 = 100$ kHz, $f_2 = 400$ kHz). The list of PD source and the corresponding nature of defects are tabulated in Table 1. The test samples were energized and left idle for one hour for it to be stabilized before commencing measurement. This idle time was needed to ensure the insulation material was able to withstand the applied voltage without breaking down during measurement. PD measurement was performed 100 times for every test sample resulting in a total of 500 PD data for all five PD sources. Every measurement data has a duration of one minute.

Two types of image format were generated from the raw PD data; 2D colour PRPD image and binary PRPD image. Samples of 2D colour PRPD images are shown in Fig. 1. The x-axis represents the phase angle of the PD occurrences, which ranges from 0° to 360°. The y-axis represents the

Table 1
PD source and the corresponding nature of defects.

PD source	Nature of defect
PD1	The insulation should be free from cavity. In this sample, a shallow 1 mm cut was made at the XLPE layer using a sharp knife.
PD2	The insulation layer should be clean prior to cable jointing. In this sample, metal particles of <1 mm size were sprinkled on the XLPE layer.
PD3	Both cable segments should be inserted into the joint at the center. However, the cable insertion was done at an off-center angle for this sample.
PD4	Cable joints should not contain voids or air pockets. A layer of air pocket was present in the insulation tape wrapping around the semiconductor edge of this sample.
PD5	The semiconductor layer should have a chamfered smooth edge. For this sample, rough edges were created at the semiconductor tip using a 6 mm rat tail file.

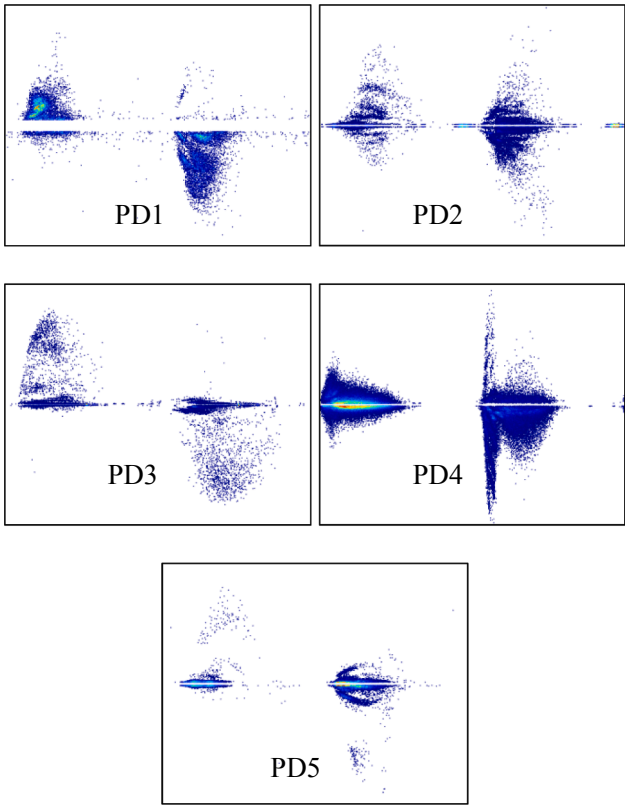

Fig. 1. Example 2D colour PRPD image of PD1 to PD5.

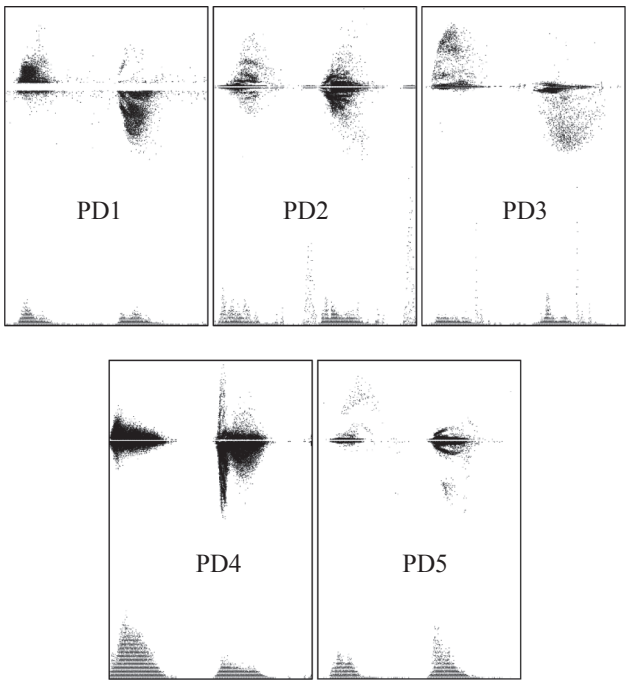

Fig. 2. Example binary PRPD image of PD1 to PD5.

Table 2
Measured PD characteristics.

Source	Duration (s)	Mean charge (pC)	Maximum charge (pC)	Pulse count
PD1	60	50.58	578.19	25,085
PD2	60	32.54	770.11	22,576
PD3	60	62.58	1110.97	130,852
PD4	60	243.40	1576.86	13,682
PD5	60	47.41	858.42	25,720
Noise	60	23.91	337.08	9198

charge magnitude of the PD in pico-Coulomb. The intensity of PD occurrence is displayed as a 'jet' type colourmap gradient where the colour changes from blue, green, yellow, and red from lowest to highest intensity. Binary PRPD image is created by stacking the phase angle vs PD charge magnitude graph on top of the phase angle vs PD intensity to form a single binary image. Samples of binary PRPD images are shown in Fig. 2.

Since the images so far consist of graphs, another two separate versions can be generated. The first is the auto scale version, where the max value of the y-axis is adjusted for every image so that all images have a similar size. The second is the fixed scale version where the maximum value of the y-axis is universal for all images. The four PD image representations used in this work are binary PRPD (auto scale) image, binary PRPD (fixed scale) image, 2D colour PRPD (auto scale) image, and 2D colour PRPD (fixed scaled) image.

2.2. Preparation of contaminated data

In order to achieve a more practical PD classification system, the PD classifier was trained using laboratory-measured data (clean data) but tested using PD data that has been overlapped with noise (contaminated data) to emulate a more realistic condition encountered on-site. For this purpose, PD noise measured from ground interference was used as the noise source [29]. After calibration, the PD detector will not detect any reading when no voltage source was applied, hence the PRPD pattern will be a flat line. However, in the event of rain, the PD detector will detect noisy signals even though when no voltage was applied. Since no changes have been made to the test object and equipment, these noisy signals must originate from ground interference. These noisy signals were recorded and used to test the noise tolerance of the PD classifier. This is a more sensible choice compared to using pseudorandom noise generated by simulation software. The characteristics of all measured PD source and noise are summarized in Table 2.

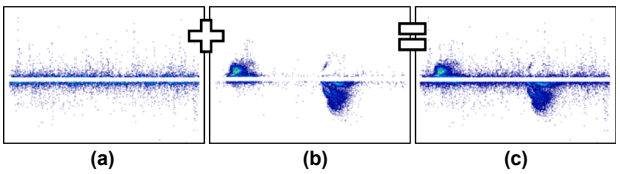

Fig. 3. Preparation of Contaminated PD data. (a) Noise PD data, (b) Clean PD data, and (c) Contaminated PD data.

A sample of the noise data and how it is overlapped to clean PD data to create contaminated PD data is shown in Fig. 3. There is a substantial difference between the clean PD data and the contaminated PD data. Therefore, the performance of a PD classification system trained using clean PD data measured in the laboratory will perform worst when being used on-site, where a certain amount of noise contamination is unavoidable. This problem can be mitigated by extracting noise invariant features that enable the classifier to recognize the PD pattern even when it has been contaminated.

3. CNN-based PD classifier

CNN is loosely based on the brain's visual cortex and how humans perceive images. The main purpose of image classification is to predict a class or the probability of classes that best illustrates the input image. For this purpose, the algorithm must be able to identify features such as ridges, edges, curves, and their structures [30]. The basic building block of CNN architecture is shown in Fig. 4. The essential layers consist of the:

- Convolutional layer: Contains numerous kernels or filters that are employed to the image in continuous positions throughout the entire image via convolution operations to produce a useful feature map for classification.
- Non-linear layer: Utilizes the ReLu (Rectified Linear Unit) activation function, which is a type of non-linear activation function. The reason for this is to prevent linearity within the system.
- Pooling layer: Downsampling is performed here to decrease the computational requirement by lowering the size of the feature maps. This also helps in improving the positional invariance of the feature map.
- Fully connected layers: Neural network receives the flattened convolutional features produced by the final convolutional layer and performs classification. The output error is described by the loss function to inform how effective the network is. The network is trained using an optimizer to minimize the loss function.

The convolutional layers of a CNN are capable of extracting intrinsic noise invariant features from an image. For example, the CNN that has been trained to recognize a dog is able to recognize a new dog of a different breed in a new location with a different background image. It is this valuable feature that makes it an excellent choice for PD classification, where the actual PD measured on-site may differ significantly from laboratory-measured data due to interference and noise contamination.

3.1. DeepLoc

DeepLoc is a CNN developed to predict eukaryotic protein subcellular localization [31]. It was trained using the UniProt database, which contains 14,000 protein data at 10 subcellular location. The DeepLoc is

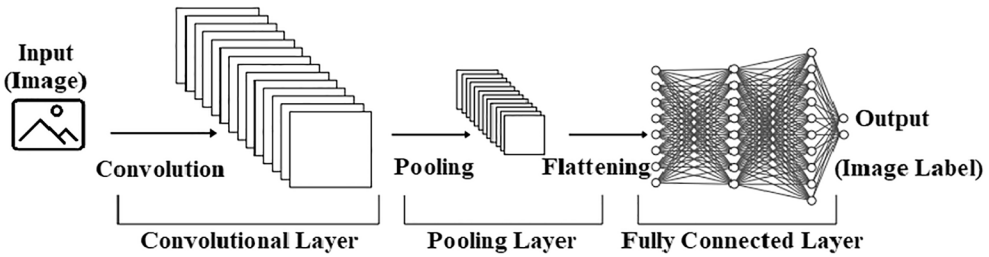

Fig. 4. Basic CNN architecture.

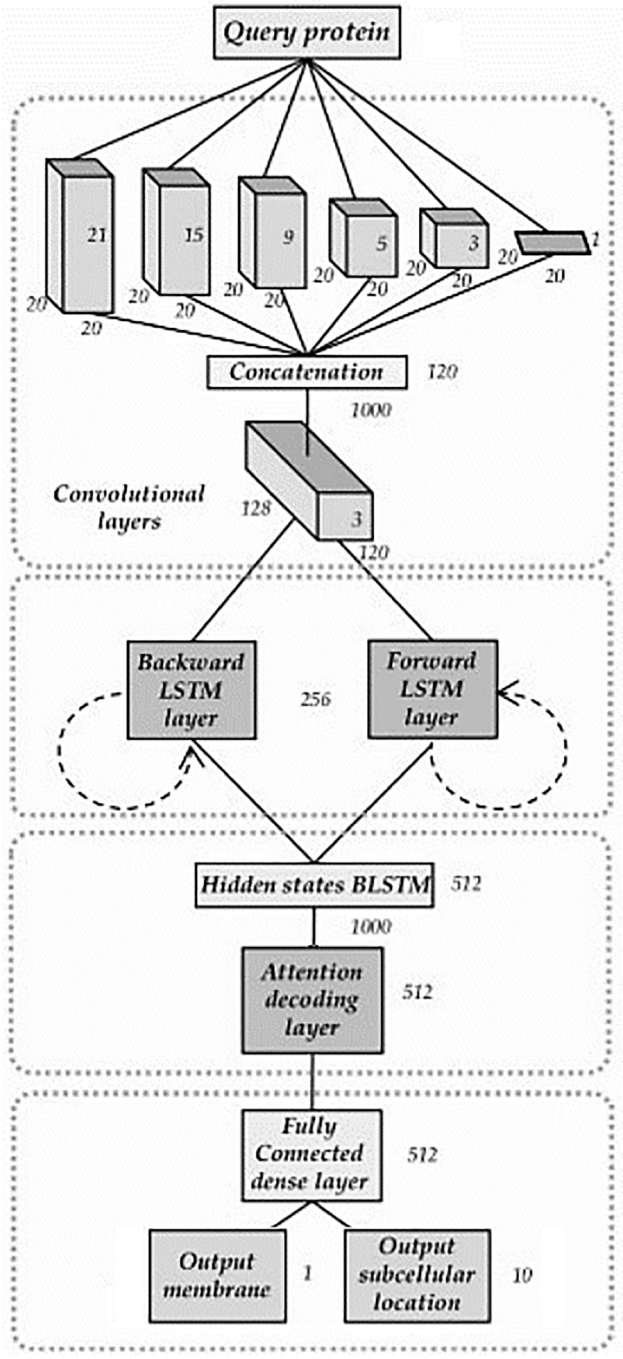

Fig. 5. DeepLoc architecture.

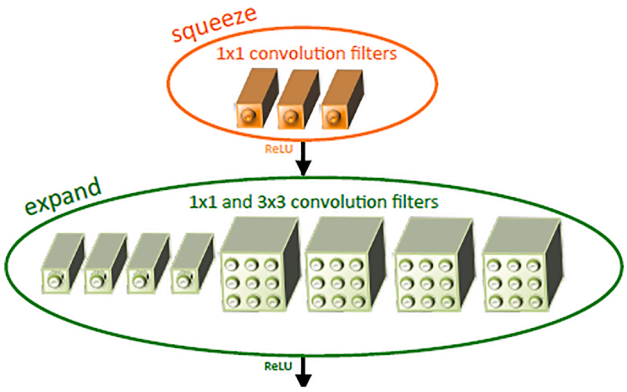

Fig. 6. Fire Module in SqueezeNet.

made up of four main layers, as shown in Fig. 5, the convolutional layer, recurrent layer, attention decoding layer, and prediction layer. The convolutional layer excerpts motif data using 120 filters of various sizes producing a 1000x128 feature map to the RNN. The RNN scans the sequence employing 256 LSTM units in two directions generating a total of 1000x512 dimensional output. The attention decoding layer employs an LSTM with 512 units through ten decoding steps. The last fully connected dense layer is comprised of a 512 neuron hidden layer and ten neuron output layer. DeepLoc can achieve up to 92% accuracy outperforming other classifiers in subcellular localization.

3.2. SqueezeNet

SqueezeNet is a type of CNN specialized for embedded applications. The network implements extreme compression in computational complexity and parameter space using squeeze layers, which is a type of channel-projection bottlenecks that employs identity-mapping shortcut connections. SqueezeNet utilizes Fire modules [32], as shown in Fig. 6, which consists of a squeeze convolution layer connected to an expand layer. SqueezeNet architecture starts with a straightforward convolution layer, linked with eight units of Fire modules, terminates with a finishing convolution layer ensued by a global average pooling layer. SqueezeNet accepts an input image of 224x224x3 and is able to extract 1000 features from the last convolution layer. SqueezeNet is able to achieve 50 times reduction in the model size and number parameters in comparison to AlexNet while surpassing the overall performance of AlexNet. It attains a top five error rate of 15.66% during the 2012 ImageNet Large Scale Visual Recognition Competition (ILSVRC).

3.3. Inception Net

Inception Net is also known as GoogLeNet. In order to efficiently expand the neural network depth, GoogLeNet uses Inception modules that are piled upon one another, hence the name Inception Net [33]. An example of an Inception module is shown in Fig. 7. This CNN consists of a chain of complex blocks of inception modules that merge spatial

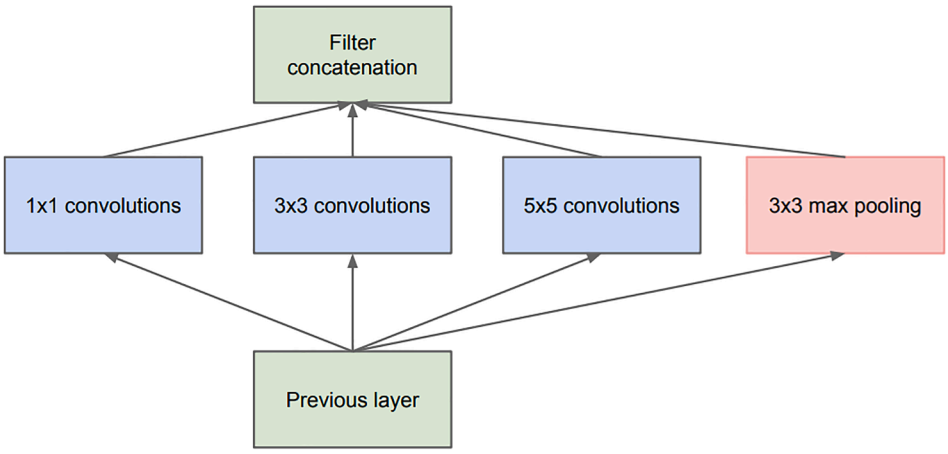

Fig. 7. Inception Module example in Inception Net.

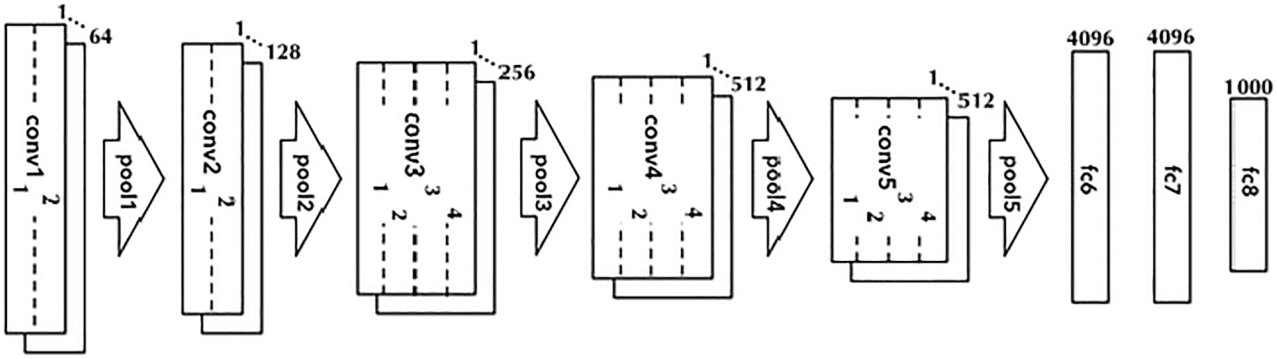

Fig. 8. VGG Net simplified architecture.

convolution, channel re-projection, and pooling operations across diverse scales in every module. The model shrinks the parameter space by decomposing spatial convolutions with bigger filter sizes (n × n) into a series of two convolutional operations with corresponding filter sizes of n × 1 and 1 × n. Supervised signals are inserted directly not just at the top conventional layer but also other layers with the objective of propagating the errors in an efficient manner. The resultant network model is a much complex and deeper network compared to AlexNet. It lowered the top five classification error to 6.656% at the 2014 ILSVRC.

3.4. VGG Net

VGG Net is one of the most well-known CNN that triumphed in the localization challenge in the 2014 ILSVRC. A simplified architecture of VGG Net is shown in Fig. 8. The number of fully connected layers and convolution layers varies from 11 to 19. As the quantity of layers was raised from 11 to 16, the top five error was decreased from 10.1% drastically. However, there was negligible improvement when the quantity of layers was raised to 19. VGG uses tiny filter size of 3x3. It was presumed that a big filter of 5x5 could be represented with two layers with a smaller filter size of 3x3 and a small amount of weights. Every

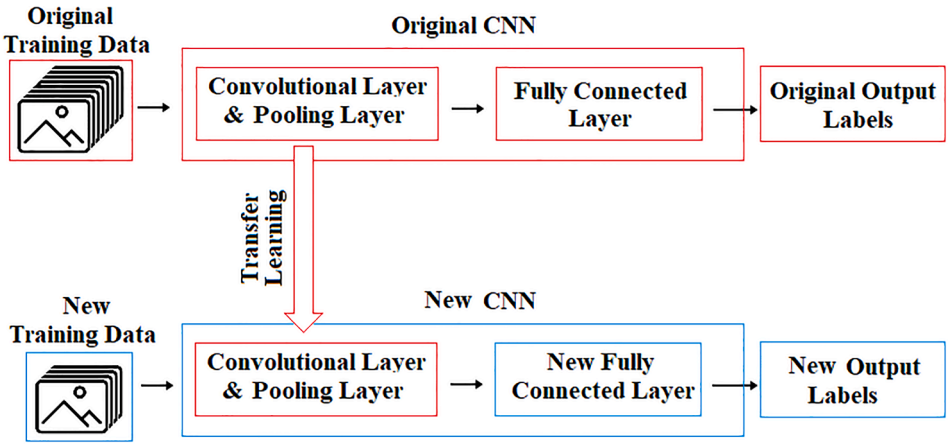

Fig. 9. Transfer learning procedure.

Table 3
Design of the new fully connected layer.

CNN	Input layer	Hidden layer	Output Layer
SqueezeNet	1000	500	5
InceptionV3	2048	1024	5
VGG16	4096	2048	5
VGG19	4096	2048	5
DeepLoc	512	256	5

sequence of convolutional layers is followed by a max pooling layer, except for the last one, which is trailed by two fully connected layers analogous to AlexNet. The two variants of VGG Net, known as VGG16 & VGG19, were used in this work.

3.5. Transfer learning implementation

Developing a new CNN from scratch requires a colossal amount of training data and computing power. For example, AlexNet CNN took six days to be trained with two GPU using ImageNet dataset, which has 15 million images of 22 thousand categories [34]. GPU capability has improved significantly in recent times, but it is difficult to obtain a PD dataset that is large enough to develop a new CNN. Therefore, a technique known as transfer learning is used, where an existing CNN is modified and repurposed to perform classification on a new PD image dataset. Transfer learning is performed by replacing CNN's last fully connected layer to a new custom neural network and retraining the CNN using a new set of training data. The convolutional layers and pooling layers of the original CNN are frozen during the training process so that only the weights of the neural network at the new fully connected layer are modified. The reason for this is to prevent the original CNN's convolutional layer from forgetting their initial unique feature extraction capability. The transfer learning procedure implemented in this work is shown in Fig. 9.

The new fully connected layer of the new CNN is a 3-layer neural network. The input layer size is determined by the size of the flattened features extracted from the convolutional layer of the original CNN. The output layer contains five neurons, one for each type of PD source. There is no available theory on how the number of neurons in the hidden layer should be chosen [35]. Upon increasing the number of hidden neurons at the hidden layer, it is found that the classification accuracy no longer steadily improves and begins to vary at a ±5% range, when the hidden layer neuron number is approximately half of the input layer neuron size. Therefore, the hidden layer neuron size was selected to be half the input layer neuron size. The new fully connected layer structure is shown in Table 3.

ReLu was used as the activation function, whereas training was done using an adaptive moment estimation (Adam) optimization algorithm [36]. Adam is an adaptive learning rate method, which calculates individual learning rates for different parameters. Adam uses approximations of the first and second moments of gradient to adjust the learning rate for each weight of the neural network. ReLu and Adam were used in this work because in terms of classification accuracy when tested against PD data overlapped with noise, it consistently outperforms other activation function such as logistic and tanh, and the stochastic gradian descend optimization algorithm.

3.6. Transfer learning

K-fold cross-validation was used to evaluate the performance of the PD classifier. This validation method was preferred compared to the holdout method at random ratio split because the variance of the result will be reduced. *K*-fold cross-validation works by splitting the dataset into *K* number of groups. The training was repeated *K* times, where every group takes a turn to act as the testing data once while all remaining groups serve as the training data. Lastly, the average accuracy

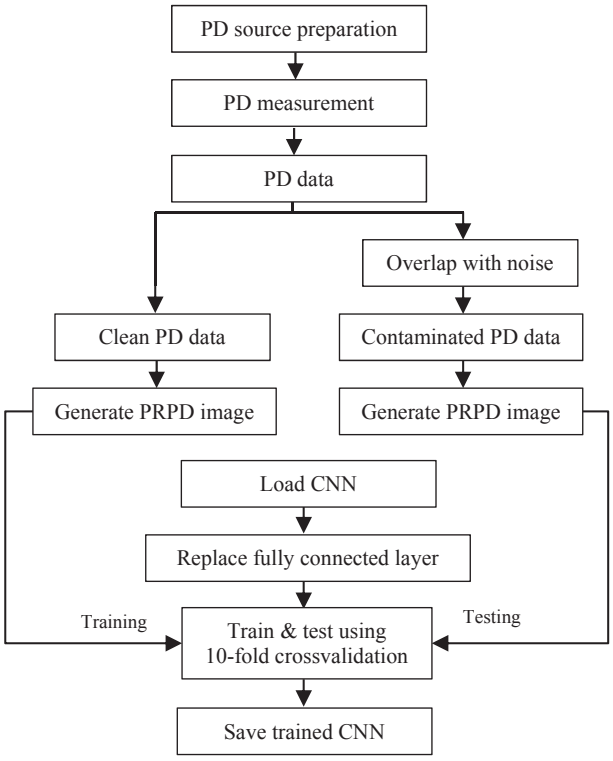

Fig. 10. Flow chart of CNN-based PD classifier preparation.

Table 4
Classification accuracy of CNN-based PD classifiers.

Input type	CNN	Clean data (%)	Contaminated data (%)
Binary PRPD (auto scale)	SqueezeNet	96.20	57.00
	InceptionV3	94.20	57.60
	VGG16	94.60	67.20
	VGG19	96.20	67.60
	DeepLoc	94.60	62.40
Binary PRPD (fixed scale)	SqueezeNet	96.40	60.40
	InceptionV3	96.60	53.40
	VGG16	94.60	56.20
	VGG19	95.20	58.00
	DeepLoc	98.40	90.80
2D colour PRPD (auto scale)	SqueezeNet	97.60	45.20
	InceptionV3	95.40	60.80
	VGG16	94.00	42.20
	VGG19	93.80	56.80
	DeepLoc	93.00	62.80
2D colour PRPD (fixed scale)	SqueezeNet	96.80	66.40
	InceptionV3	93.20	78.20
	VGG16	94.60	62.60
	VGG19	94.80	57.80
	DeepLoc	96.20	63.00

across all *K* trials was calculated. By increasing the value of *K*, the result variance will decrease, but the training time will increase. In this work, 10-fold cross-validation was used. The entire process of preparing the CNN-based PD classifier is summarized in a flow chart in Fig. 10.

In order to evaluate the performance of the PD classifier during noise contamination, a slight modification was done during the 10-fold cross-validation procedure. During each round of the training in the modified 10-fold cross-validation procedure, the training data was left unmodified, but the testing data was overlapped with noise. This mimics on-site usage of the PD classifier, where it will be trained using clean data, but tested with noise-contaminated data that it has not seen before during

Table 5
Performance comparison on the same PD dataset and noise source.

PD Classifier	Input	Clean data (%)	Contaminated data (%)
ML: ANN	PCA	89.00	73.30
	Fractal	85.44	26.50
	Statistical	91.57	52.40
ML: SVM	PCA	93.35	73.90
	Fractal	91.93	57.10
	Statistical	98.67	47.10
CNN: InceptionV3	2D colour PRPD (fixed scale)	93.20	78.20
CNN: DeepLoc	Binary PRPD (fixed scale)	98.40	90.80

training.

4. Results and discussion

In order to repurpose the CNN for other applications, the input data of the new application such as PD classification must be represented in image format. The type of image format representation will affect the final classification accuracy as every CNN is trained to extract different features that are relevant for their original intended tasks. Therefore, in this work, four types of input features were investigated. The classification accuracy of the CNN-based PD classifiers after performing transfer learning using all four types of input image format is shown in Table 4. Regular 10-fold cross-validation was used to observe how well the model performed using clean data measured from the laboratory. Modified 10-fold cross-validation was done to test the robustness of the model in handling noise-contaminated data.

4.1. Performance analysis

By observing the obtained results, it is clear that all five CNNs are able to perform very well regardless of the input data type for clean PD data classification. This shows that it is feasible to develop a PD classification system using PRPD images as the input. The CNNs in the current work achieved classification accuracy between 93.20% and 98.4% for clean PD data. Such performance is comparable to traditional machine learning algorithms for classifying clean PD data.

When dealing with contaminated data, all CNN suffer classification accuracy reduction as expected. From the results, there is no clear trend of whether which particular CNN or input type is superior to the other. For example, SqueezeNet performed better when using 2D colour PRPD fixed scale image (66.40%) but worse when using the auto scale version of 2D colour PRPD (45.20%). InceptionV3 generally performs better when using colour image (60.80% & 78.20%) instead of binary image (57.60 & 53.40%). VGG16 & VGG19 performed better when using fixed scale version of 2D colour PRPD image but worst when using fixed scale version of binary PRPD image. On the other hand, DeepLoc, performed better when using fixed scaled version of both input types.

The best performing CNN and image type combination were DeepLoc with fixed scale binary PRPD image where it retained an accuracy of 90.80% under noise contamination. InceptionV3 was the second-best, with a performance of 78.20% when using fixed scale 2D colour PRPD under noise contamination.

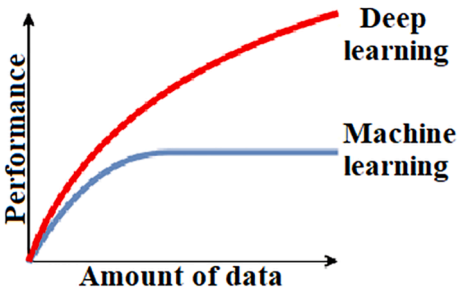

Fig. 11. Deep learning vs machine learning [40].

4.2. Comparison with machine learning-based PD classifier

For a direct apples-to-apples comparison, the results of the current work were compared to a previous work which used the same PD dataset and noise source. In [28], traditional machine learning methods such as shallow ANN and SVM were used to classify the same five PD sources that were overlapped with noise using feature extraction methods such as statistical parameters, fractal features, and principal component analysis (PCA). A summary of the performance achieved by traditional machine learning and CNN is shown in Table 5. The overall best-performing machine learning classifier is SVM with PCA features where it obtained 93.35% for clean data and 73.90% for contaminated data. CNN-based PD classifier, DeepLoc managed to outperform it in both clean and contaminated data where it obtained 98.40% and 90.80%, respectively. The superiority of CNN-based classifier is evident for its 16.90% better classification accuracy when dealing with contaminated data.

4.3. Comparison with other works

Noise analysis of PD classification performed by other researchers is summarized in Table 6. A direct comparison with the current work is not possible due to different PD source type, number of PD source and the nature of noise used. Nevertheless, it can still serve as a rough estimation of how the current work compares to others.

As the number of classes increases, it would be tougher for the classifier to achieve good accuracy because the likelihood of a wrong prediction is higher. Therefore, the current work presented a more challenging task for the PD classifier as it had to classify five types of PD sources compared to three to five PD sources used in other works. Apart from that, the current work used real measured noise of 44% instead of pseudorandom simulated noise ranging from 15% to 40% in terms of charge magnitude. Since the noise source used varies in different works, only the maximum noise level of each work was included in the comparison.

Table 6
Classification accuracy of other noise-related PD classifiers.

Ref.	PD classifier	Noise source	PD source	Clean data (%)	Contaminated data (%)
[37]	ANN + 3D PRPD	40% simulated noise	4 types, current transformer	100	70.5
[18]	ANN + statistical features	30% simulated noise	3 types, transformer	96.3	80.0
[38]	ANN + Hilbert spectrum	15% simulated noise	4 types, XLPE cable	100	76.7
[7]	ANN + fractal features	30% simulated noise	5 types, epoxy material	100	83.8
[39]	ANN + fractal features	30% simulated noise	4 types, XLPE cable	100	69.9
Current	InceptionV3 + PD images	44% measured noise	5 types, XLPE cable joints	93.20	78.20
Current	DeepLoc + PD images	44% measured noise	5 types, XLPE cable joints	98.40	90.80

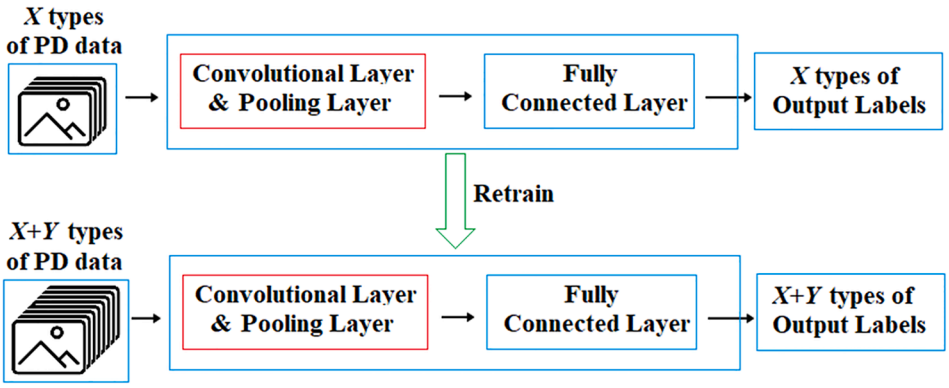

Fig. 12. Expanding the capability of CNN-based PD classifier.

Despite the more challenging situation for the CNN-based PD classifier, the deterioration in accuracy is lower compared to other works when dealing with contaminated data. By using ANN trained with manually extracted features from the PD data, the classification accuracy of clean data is above 95% for most cases where some even have perfect accuracy. However, it is evident that traditional machine learning with manual feature extraction suffers more deterioration when noise is present. When dealing with noise-contaminated data, classification accuracy in other works suffer a reduction ranging from 16.2% to 31.1%. On the other hand, in current work, the best noise invariant CNN-based classifier is DeepLoc CNN, with only a 7.6% reduction in accuracy (from 98.40% to 90.80%) when dealing with contaminated data.

4.4. *Further advantages of CNN-based PD classification system*

Apart from the highly invariant automatic feature extraction, deep learning methods such as CNN has an added advantage of easy scalability as they are meant to be trained with a large amount of data that has many output classes. As shown in Fig. 11, unlike traditional machine learning, the performance of deep learning will plateau at a higher level when more data is added. This is evident, as shown during the ILSVRC, where deep learning vastly outperforms machine learning when a large amount of data was involved [34].

It is also feasible to add new PD source recognition capability to the existing CNN-based PD. For example, it is possible to train the CNN-based PD classifier to learn to recognize new types of PD patterns from when PD data from new insulation defect type is added into the database. As shown in Fig. 12, by adding more neurons to the output layer and retraining the CNN with the original PD dataset plus the new groups of PD data from the new sources, the CNN-based PD classifier will be able to recognize more types of PD patterns.

5. Conclusion

The feasibility of using a CNN-based PD classification system had been successfully investigated. CNN-based PD classifier was able to achieve excellent accuracy for classifying clean PD patterns from five types of PD sources with an accuracy of above 93%. When being tested with contaminated PD data, CNN-based PD classifier based on DeepLoc was able to outperform traditional machine learning by a wide margin, which is 90.80% against 73.90% (16.90% improvement). Thus, it has been shown that the CNN-based PD classification system was more robust and has higher noise tolerance than traditional machine learning-based systems. The easy scalability of CNN also makes it a more effective option compared to traditional machine learning models. Currently, only five types of PD source were tested due to the limited available resources. Although it is theoretically possible to expand the capability of the proposed PD classification to recognize additional PD sources, it has not been implemented yet. This will be explored in a future work if a large-scale public PD dataset is made available.

CRediT authorship contribution statement

Wong Jee Keen Raymond: Funding acquisition, Formal analysis, Writing - review & editing, Supervision. **Chong Wan Xin:** Methodology, Software, Writing - original draft, Validation, Investigation, Data curation. **Lai Weng Kin:** Project administration, Resources. **Hazlee Azil Illias:** Conceptualization, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported in part by TARUC Internal Research Grant (UC/I/G2018-00026), University of Malaya Faculty Research Grant (GPF077A-2018) and Nvidia Corporation sponsored GPU.

References

[1] E.T. Iorkyase, C. Tachtatzis, P. Lazaridis, D. Upton, B. Saeed, I. Glover, et al., Improving RF-based partial discharge localization via machine learning ensemble method, IEEE Trans. Power Delivery (2019) 13.
[2] M. Homaei, S.M. Moosavian, H.A. Illias, Partial discharge localization in power transformers using Neuro-Fuzzy technique, IEEE Trans. Power Delivery 29 (2014) 2066–2076.
[3] H. Tao, M.T.C. Fang, Detection and classification of partial discharge using a feature decomposition-based modular neural network, IEEE Trans. Instrum. Meas. 50 (2001) 1349–1354.
[4] K. Firuzi, M. Vakilian, B.T. Phung, T.R. Blackburn, Partial discharges pattern recognition of transformer defect model by LBP & HOG features, IEEE Trans. Power Delivery 34 (2019) pp. 542–550.
[5] E. Gulski, Computer-aided measurement of partial discharges in HV equipment, IEEE Trans. Electr. Insul. 28 (1993) 969–983.
[6] V.P. Darabad, M. Vakilian, B.T. Phung, T.R. Blackburn, An efficient diagnosis method for data mining on single PD pulses of transformer insulation defect models, IEEE Trans. Dielectrics Electrical Insul. 20 (2013) pp. 2061–2072.
[7] H.-C. Chen, F.-C. Gu, Pattern recognition with cerebellar model articulation controller and fractal features on partial discharges, Expert Syst. with Appl. 39 (2012/06/01/ 2012) pp. 6575–6584.
[8] T. Yuming, Z.D. Wang, P.A. Crossley, Partial discharge pattern recognition based on 2-D wavelet transform and neural network techniques, in: Power Engineering Society Summer Meeting, vol.1, 2002 IEEE, 2002, pp. 411–416.
[9] F.-C. Gu, H.-C. Chen, B.-Y. Chen, A fractional fourier transform-based approach for gas-insulated switchgear partial discharge recognition, J. Electric. Eng. Technol. July 08 2019.
[10] F. Gu, H. Chen, M. Chao, Application of improved hilbert-huang transform to partial discharge defect model recognition of power cables, Appl. Sci. 7 (2017) 1021.
[11] J. Seo, H. Ma, T. Saha, Probabilistic wavelet transform for partial discharge measurement of transformer, IEEE Trans. Dielectr. Electr. Insul. 22 (2015) 1105–1117.

[12] Y. Xu, Y. Qian, F. Yang, Z. Li, G. Sheng, X. Jiang, DC cable feature extraction based on the PD image in the non-subsampled contourlet transform domain, IEEE Trans. Dielectr. Electr. Insul. 25 (2018) 533–540.
[13] R. Yao, M. Hui, J. Li, L. Bai, Q. Wu, A new discharge pattern for the characterization and identification of insulation defects in GIS, Energies 11 (2018) 971.
[14] C. Mazzetti, F.M.F. Mascioli, F. Baldini, M. Panella, R. Risica, R. Bartnikas, Partial discharge pattern recognition by neuro-fuzzy networks in heat-shrinkable joints and terminations of XLPE insulated distribution cables, IEEE Trans. Power Delivery 21 (2006) pp. 1035–1044.
[15] X. Peng, J. Li, G. Wang, Y. Wu, L. Li, Z. Li, et al., Random forest based optimal feature selection for partial discharge pattern recognition in HV cables, IEEE Trans. Power Delivery 34 (2019) pp. 1-1.
[16] W.J.K. Raymond, H.A. Illias, A.H.A. Bakar, H. Mokhlis, Partial discharge classifications: Review of recent progress, Measurement 68 (2015) 164–181, 2015/05/01/.
[17] A.A. Mas'ud, R. Albarracín, J.A. Ardila-Rey, F. Muhammad-Sukki, H.A. Illias, N.A. Bani, et al., Artificial neural network application for partial discharge recognition: survey and future directions, Energies 9 (2016) p. 574.
[18] C.-C. Kuo, Artificial identification system for transformer insulation aging, Expert Syst. Appl. 37 (2010.) 4190–4197, 2010/06/01/.
[19] H. Song, J. Dai, G. Sheng, X. Jiang, GIS partial discharge pattern recognition via deep convolutional neural network under complex data source, IEEE Trans. Dielectr. Electr. Insul. 25 (2018) 678–685.
[20] W. Jee Keen Raymond, H.A. Illias, A.H. Abu Bakar, Classification of partial discharge measured under different levels of noise contamination, PLOS ONE 12 (2017) p. e0170111.
[21] V.M. Catterson, B. Sheng, Deep neural networks for understanding and diagnosing partial discharge data, in: 2015 IEEE Electrical Insulation Conference (EIC), 2015, pp. 218–221.
[22] M.T. Nguyen, V.-H. Nguyen, S.-J. Yun, Y.-H. Kim, Recurrent neural network for partial discharge diagnosis in gas-insulated switchgear, Energies 11 (2018) p. 1202, 05/09 2018.
[23] J. Tang, M. Jin, F. Zeng, X. Zhang, R. Huang, Assessment of PD severity in gas-insulated switchgear with an SSAE, IET Sci. Meas. Technol. 11 (2017) 423–430.
[24] X. Wang, H. Huang, Y. Hu, Y. Yang, Partial discharge pattern recognition with data augmentation based on generative adversarial networks, in: 2018 Condition Monitoring and Diagnosis (CMD), 2018, pp. 1–4.
[25] Q. Zhang, J. Lin, H. Song, G. Sheng, Fault Identification Based on PD Ultrasonic Signal Using RNN, DNN and CNN, 2018.
[26] N. Duan, E.M. Stewart, Deep-learning-based power distribution network switch action identification leveraging dynamic features of distributed energy resources, IET Gener. Transm. Distrib. 13 (2019) 3139–3147.
[27] J. Wang, Z. Mo, H. Zhang, Q. Miao, A deep learning method for bearing fault diagnosis based on time-frequency image, IEEE Access 7 (2019) 42373–42383.
[28] M. Xia, T. Li, L. Xu, L. Liu, C.W.d. Silva, Fault diagnosis for rotating machinery using multiple sensors and convolutional neural networks, IEEE/ASME Trans. Mechatronics 23 (2018) pp. 101–110.
[29] W.J.K. Raymond, H.A. Illias, A.H.A. Bakar, High noise tolerance feature extraction for partial discharge classification in XLPE cable joints, IEEE Trans. Dielectr. Electr. Insul. 24 (2017) 66–74.
[30] Y. LeCun, K. Kavukcuoglu, C. Farabet, Convolutional networks and applications in vision, in: 2010 IEEE International Symposium on Circuits and Systems (ISCAS), 2010, pp. 253–256.
[31] J.J. Almagro Armenteros, C.K. Sønderby, S.K. Sønderby, H. Nielsen, O. Winther, DeepLoc: prediction of protein subcellular localization using deep learning, Bioinformatics 33 (2017) 3387–3395.
[32] F. Iandola, S. Han, M. Moskewicz, K. Ashraf, W. Dally, K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size," 02/23 2016.
[33] C. Szegedy, L. Wei, J. Yangqing, P. Sermanet, S. Reed, D. Anguelov, et al., Going deeper with convolutions, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9.
[34] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in: presented at the Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, Lake Tahoe, Nevada, 2012.
[35] K.X. Lai, B.T. Phung, T.R. Blackburn, Application of data mining on partial discharge part I: predictive modelling classification, in: IEEE Transactions on Dielectrics and Electrical Insulation, vol. 17, 2010, pp. 846–854.
[36] D. Kingma, J. Ba, Adam: A method for stochastic optimization, International Conference on Learning Representations, 12/22 2014.
[37] C. Hong-Chan, K. Ying-Piao, L. Chun-Yao, L. Han-Wei, A partial discharge based defect-diagnosis system for cast-resin current transformers, in: 39th International Universities Power Engineering Conference, Vol. 1, 2004. UPEC 2004, 2004, pp. 233-237.
[38] F.C. Gu, H.C. Chang, F.H. Chen, C.C. Kuo, C.H. Hsu, Application of the Hilbert-Huang transform with fractal feature enhancement on partial discharge recognition of power cable joints, Sci., Meas. Technol., IET 6 (2012) 440–448.
[39] H.C. Chen, Fractal features-based pattern recognition of partial discharge in XLPE power cables using extension method, Gener., Trans. Distrib, IET 6 (2012) 1096–1103.
[40] A. Ng, Machine Learning Yearning: Technical Strategy for AI Engineers, In the Era of Deep Learning, deeplearning.ai, p. 11, 2018.